Engineering Research Institute

UNIVERSITY OF MASSACHUSETTS

AMHERST, MASSACHUSETTS

19

MINIMAX DESIGN OF LOW SENSITIVITY
FILTERS FOR STATE ESTIMATION

by

Joseph A. D'Appolito

December 1969

Technical Report No. 4

MINIMAX DESIGN OF LOW SENSITIVITY FILTERS FOR STATE ESTIMATION

by

Joseph A. D'Appolito

Prepared under Grant NGR 22-010-012

for the

National Aeronautics and Space Administration

ACCEPTANCE

This report is Dr. Joseph A. D'Appolito's doctoral dissertation which was submitted to the University of Massachusetts for partial fulfillment of the requirements for the degree of Doctor of Philosophy. Reproduction in whole or part is permitted for any purpose of the United States Government.

Charles E. Hutchinson

Charles E. Hutchinson
Co-principal Investigator
Grant NGR 22-010-012
University of Massachusetts

Minimax Design of Low Sensitivity Filters

for State Estimation. (September 1969)

Joseph A. D'Appolito, B.E.E., Rensselaer Polytechnic Institute

M.S.E.E. and E.E., Massachusetts Institute of Technology

Directed by: Prof. Charles E. Hutchinson

Realization of the optimal (Kalman) filter for estimating the state of a linear system from noisy measurements requires exact knowledge of plant dynamics and plant and measurement noise statistics. The question of how one designs a state estimator in the presence of large uncertainties in these parameters naturally arises. This thesis proposes minimax design criteria for this purpose and explores the theoretical and computational aspects of the resulting minimax problems.

A linear estimator identical in form to the Kalman filter is chosen a priori. Three useful performance measures for this filter are its total mean square estimation error (S_1), and the deviation of this error from the optimum (minimum) estimation error in either an absolute (S_2) or relative (S_3) sense. These performance measures are a function of the uncertain parameters and a set of unspecified filter feedback and feedforward gains. If a particular performance measure is first maximized over the uncertain parameter set and then minimized with respect to the adjustable gains, a filter is obtained which yields a least upper

bound on the performance measure regardless of the exact value of the uncertain parameters.

Minimax filter design for time-invariant plants with constant, but uncertain plant and measurement noise statistics is fully explored. First, for the S_1 performance measure, it is shown that min-max equals max-min. This result allows one to replace the minimax problem with a simple maximization of the optimal performance measure over the uncertain parameter set. The S_1 filter is then simply the Kalman filter for the maximizing noise statistics. Many properties of the required maximization for the infinite time case are developed and several examples given.

Next the S_2 and S_3 filters for time invariant plants with constant, but uncertain noise statistics are shown to be unique. It is established that the S_2 and S_3 filter gains are optimal for at least one point in the set of uncertain parameters. Unfortunately min-max does not equal max-min for the S_2 and S_3 performance measures so one is forced to solve the complete minimax problem. The convexity of the S_2 and S_3 performance measures in the uncertain parameter set is established, however, and used to show that the maximum of these performance measures is attained over a known finite set of points.

Finally, the minimax filtering problem for plant and measurement noise statistics with uncertain or arbitrary time variation is investigated in detail. It is shown that the S_1 performance measure can be found by maximizing the optimal filter performance measure over the set of admissable trajectories in the uncertain parameter space and several properties of this maximization are developed.

A general approach to the design of linear filters for state estimation when large uncertainties in plant and measurement parameters are present is given. The case of uncertain noise statistics is fully explored and several examples are presented to illustrate the utility and practicality of the approach.

ACKNOWLEDGMENT

Partial support for this effort was obtained from the National Aeronautics and Space Administration under Grant NGR 22-010-012.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

The list given below contains symbols in general use throughout the text. Additional specialized symbols of limited usage, such as those appearing in examples, are defined as they are encountered.

A	domain of α
B	domain of β
C	composition of V_F and V_K
C_n	a convex polyhedral cone in V_K
E	ensemble expectation operator
E_n	euclidean n space
F	$n \times n$ matrix of plant dynamics
G	$n \times m$ plant gain matrix
H	$p \times n$ measurement matrix, also the scalar Hamiltonian
H_o	minimum value of Hamiltonian
J_M	scalar performance index of arbitrary filter
$\bar{J}_M$	infinite time J_M
$\bar{J}_M^i$	basis functional of J_M
J_o	optimum (minimum) J_M
$\bar{J}_o$	infinite time J_o

$\mathcal{K}_o$	a set of K_o matrices
$\overline{\mathcal{K}}_o$	a set of $\overline{K}_o$ matrices
K	$p \times n$ gain matrix of arbitrary filter
K_o	optimum K
$\overline{K}_o$	infinite time K_o
K^*	minimax K
$\hat{K}$	minimax sensitivity K
ΔK	an increment in V_K
M	$n \times n$ covariance matrix for arbitrary estimate
$\overline{M}$	infinite time M
m	dimension of u
n	dimension of x
P	optimum M
P_o	initial value of P
$\overline{P}$	infinite time P
P_{ij}	partial of P with respect to q_{ij} or r_{ij}
$\overline{P}_{ij}$	infinite time P_{ij}
Q	$m \times m$ covariance matrix for u
Q_{ij}	partial of Q with respect to q_{ij}
q_{ij}	ijth element of Q
q_{max}	upper bound on trace Q
q_{min}	lower bound on trace Q

R	$p \times p$ covariance matrix for w
R_{ij}	partial of R with respect to r_{ij}
r_{ij}	ijth element of R
r_{max}	upper bound on trace R
r_{min}	lower bound on trace R
S	open half space in V_K
S^A	absolute performance sensitivity
S^R	relative performance sensitivity
S_1	minimax value of J_M
S_2	minimax value of S^A
S_3	minimax value of S^R
s	laplace transform frequency variable
u	m vector of plant noise
V	composition of V_Q and V_R
V_F	a set of F matrices
V_K	a set of K matrices
V_Q	set of uncertain Q
V_R	set of uncertain R
v	elements of V
v_i	extreme point of V
v_T^*	maximizing v at time T

W	$n \times n$ positive definite weighting matrix
w	p vector of measurement noise
X	a convex set
X_E	the set of extreme points of X
x	$n \times 1$ plant state vector
$\bar{x}$	arbitrary estimate of x
$\hat{x}$	optimum estimate of x
z	$p \times 1$ measurement vector
α	vector of unknown parameters
α_i	set of indexed constants
β	vector of adjustable parameters
β^*	minimax β
γ	a constant
γ_i	a set of indexed constants
Λ	$n \times n$ costate matrix
ρ	correlation coefficient
Φ_K	$n \times n$ state transition matrix for filter with gain K
Ψ	$n \times n$ transition matrix for costate equation
$\nabla_K(\cdot)$	gradient of $(\cdot)$ with respect to K

CHAPTER I

INTRODUCTION

1.1 Background

In recent years the Kalman [1] or Kalman-Bucy [2] filter has gained wide acceptance as an estimator for the state vector of a linear dynamical system excited by random inputs. This filter is optimum in the sense that it generates the minimum variance unbiased estimate of the system state vector from noisy measurements of the output. The ability of this filter to handle time-varying systems and its ready realization on a digital computer have contributed greatly to its popularity. A major drawback to the use of the Kalman filter, however, lies in the fact that its realization requires exact knowledge of the system dynamics and the covariance matrices of the system input and measurement noises. These parameters are rarely known exactly. Usually they are known only approximately, or equivalently, one can state only the probable range in which they lie. The question of how to design a filter in the face of these uncertainties immediately arises.

To date, two approaches to this problem have received serious attention in the literature. In the first [3], one attempts to estimate the unknown parameters along with the state vector. In the case of unknown dynamic parameters this is usually accomplished by assuming that the unknown parameters satisfy some set of differential equations and appending these equations to the plant equations. Unfortunately, the assumed parameter dynamics can rarely be justified. Furthermore, the augmented system of equations is invariably nonlinear. Thus not only is one forced to estimate additional variables, but unless suitable linearizations are made, the Kalman filter is no longer even applicable. These linearizations usually place serious limitations on the allowable range of parameter uncertainty. Maximum likelihood techniques have been developed for estimating both dynamic and statistical parameters [4, 5, 6], but these estimators are again nonlinear, and more importantly, non-recursive in structure. This latter property makes the practical application of these estimators to systems of any reasonable size computationally prohibitive. Furthermore, the error performance of these estimators on finite data records is unknown. Finally, as in all cases of nonlinear estimation, the question of whether or not the parameters of interest are even observable is largely unanswered.

In the second approach [7], one designs a Kalman filter for some nominal value of the unknown parameters. The error performance of this nominal filter is then compared with the optimal error over the entire assumed range of the unknown parameters. This procedure is repeated for several nominal values until a filter is obtained in which the departure from optimality is acceptable. Though this approach often leads to an engineeringly useful filter, it requires extensive computer simulation and in the end one is never certain that a better filter does not exist.

In the second approach, by comparing the error performance of a nominal filter against that of the optimal filter, one is inherently invoking the concept of sensitivity. The procedure as it stands, however, is basically analytical and the concept of sensitivity is only qualitatively defined. A procedure for directly synthesizing low sensitivity filters is clearly desirable. In order to accomplish this, one must first quantify the concept of sensitivity in a manner appropriate to the filtering problem and then derive algorithms for optimizing the sensitivity measure, thereby realizing the design of the optimally insensitive filter. This in general terms is the goal of the research reported herein.

The basic ground rule for the conduct of this research has been that all design techniques evolved must be readily applicable and must lead to filter mechanizations of no greater complexity then that of the Kalman filter.

1.2 Problem Statement

Consider the linear plant described by the vector-matrix differential equation

$$\dot{x}(t) = F(\alpha_1)\,x(t) + G(t)\,u(t) \tag{1-1}$$

with noisy measurement

$$z(t) = H(t)\,x(t) + w(t)\,; \qquad t_o \le t \le T \tag{1-2}$$

In the above equations x(t), u(t), and z(t) are column vectors of n, $m \le n$ and p dimension respectively. F, G, and H are matrices of appropriate dimension. The vectors u(t) and w(t) are uncorrelated zero mean Gaussian white noise processes such that

$$\mathrm{Cov}\,[u(t)] = E\left\{u(t)\,u^T(t)\right\} = Q(\alpha_2)\,\delta(t-\tau) \tag{1-3}$$

and

$$\mathrm{Cov}\,[w(t)] = E\left\{w(t)\,w^T(t)\right\} = R(\alpha_3)\,\delta(t-\tau) \tag{1-4}$$

where Q is a non-negative definite $m \times m$ matrix and R is a positive definite $p \times p$ matrix.

The α_i are vectors representing uncertain parameters which are known only to lie in compact sets A_i. In the most general case both α_i and A_i may be time varying. Uncertainties in the H and G matrices are considered adequately reflected by uncertainties in the Q and R matrices. For expository convenience the combined parameter vector $\alpha^T = \left(\alpha_1^T \mid \alpha_2^T \mid \alpha_3^T\right)$ is defined.

If α is known and the optimum estimate of x(t) is desired, one simply builds the Kalman filter for that value of α. This filter has the form [2]:

$$\dot{\hat{x}}(t) = F(\alpha_1)\,\hat{x}(t) + K_o(t)[z(t) - H(t)\,\hat{x}(t)] \tag{1-5}$$

where $\hat{x}(t)$ is that estimate of $x(t)$ in the class of all unbiased estimators $\overline{x}(t)$ of the state vector which minimizes the scalar cost function

$$J(t) = E\left\{\|\,\overline{x}(t) - x(t)\,\|^2\right\} \tag{1-6}$$

given the observation record $z(t)$ from t_o to time t. $J(t)$ is simply the mean square estimation error at time t. The optimum filter gain is given by

$$K_o(t) = P(t)\,H(t)\,R^{-1}(\alpha_3) \tag{1-7}$$

$P(t)$ is the covariance of the optimum estimate defined as

$$P(t) = E\left\{[x(t) - \hat{x}(t)][x(t) - \hat{x}(t)]^T\right\} \tag{1-8}$$

$P(t)$ is obtained from the solution to the matrix Riccati equation

$$\begin{aligned}\dot{P}(t) = {} & F(\alpha_1)\,P(t) + P(t)\,F^T(\alpha_1) + G(t)\,Q(\alpha_2)\,G^T(t) \\ & - P(t)\,H^T(t)\,R^{-1}(\alpha_3)\,H(t)\,P(t)\end{aligned} \tag{1-9}$$

The initial conditions for (1-5) and (1-9) are respectively:

$$\hat{x}(t_o) = 0 \tag{1-10}$$

and

$$P(t_o) = P_o = E\{x_o x_o^T\} \tag{1-11}$$

Using the definition of P(t), eq. (1-8), the optimum (minimum) value of J(t) is

$$J_o(t) = \mathrm{tr}\,[P(t)] = \mathrm{tr}\,[P(\alpha,\ P_o,\ t_o,\ t)] \tag{1-12}$$

Note that $J_o(t)$ is a function of P_o which may also be uncertain.

As indicated in Section 1.1, the major advantages of the Kalman filter, aside from its optimality, are that it is recursive and therefore readily mechanized. It is desirable to retain these latter advantages in any optimally insensitive filter. Lacking exact knowledge of α, one therefore selects a filter identical in form to that of the Kalman filter to estimate x(t). Now, however, the feedback and feedforward gains will be adjusted in such a manner as to satisfy an appropriate sensitivity criterion. Specifically, the filter has the form

$$\dot{\bar{x}}(t) = F(\beta_1)\bar{x}(t) + K(\beta_2)[z(t) - H(t)\bar{x}(t)] \tag{1-13}$$

Assume for the moment that the filter (1-13) is uniformly asymptotically stable and that the vector $\beta^T = \left(\beta_1^T \mid \beta_2^T\right)$ belongs to a compact set B of sufficient range and dimension to generate the Kalman filter for every $\alpha \epsilon A$. The manner in which these requirements are enforced will become apparent in the sequel.

Let

$$M(t) = \mathrm{Cov}\,[x(t) - \bar{x}(t)] = E\left\{[x(t) - \bar{x}(t)][x(t) - \bar{x}(t)]^T\right\} \qquad (1\text{-}14)$$

and note that

$$M(T) = M(\alpha, \beta, P_o, t_o, T) \qquad (1\text{-}15)$$

The mean square estimation error for the filter (1-13) is then

$$J_M(T) = \mathrm{tr}\, M(T) \qquad (1\text{-}16)$$

The value of J_M is a function of both the unknown system parameters α and the adjustable filter parameters β. From the definition of J_o it is clear that

$$J_M(\alpha, \beta, P_o, t_o, T) \geq J_o(\alpha, P_o, t_o, T) \geq 0; \qquad (1\text{-}17)$$

$$\forall \alpha \in A \quad \text{and} \quad \forall \beta \in B$$

J_M is certainly a measure of the performance of the filter (1-13). Since for a given α, J_o is known, it also seems appropriate to measure the performance sensitivity of filter (1-13) in terms of its absolute or relative departure from optimality. These performance sensitivity measures take the form

$$S^A(T) = J_M(\alpha, \beta, P_o, t_o, T) - J_o(\alpha, P_o, t_o, T) \qquad (1\text{-}18)$$

$$S^R(T) = \frac{\left[J_M(\alpha, \beta, P_o, t_o, T) - J_o(\alpha, P_o, t_o, T)\right]}{J_o(\alpha, P_o, t_o, T)} \tag{1-19}$$

A rational criterion for determining the design of filter (1-13), that is, for selecting the value of β is now desired. Since α is uncertain and β alone is available for selection by the designer it seems most appropriate to select β so as to satisfy one of the following criteria:

$$S_1(P_o, t_o, T) = \min_{\beta \epsilon B} \max_{\alpha \epsilon A} J_M(\alpha, \beta, P_o, t_o, T) \tag{1-20}$$

$$S_2(P_o, t_o, T) = \min_{\beta \epsilon B} \max_{\alpha \epsilon A} S^A(\alpha, \beta, P_o, t_o, T) \tag{1-21}$$

$$S_3(P_o, t_o, T) = \min_{\beta \epsilon B} \max_{\alpha \epsilon A} S^R(\alpha, \beta, P_o, t_o, T) \tag{1-22}$$

All three criteria are minimax in nature. The S_1 criterion simply minimizes the maximum value of J_M over the uncertain parameter space A. Letting β^* denote a value of β which yields the minimax one has

$$\max_{\alpha \epsilon A} J_M(\alpha, \beta) \geq \max_{\alpha \epsilon A} J_M(\alpha, \beta^*) \geq J(\alpha, \beta^*) \tag{1-23}$$

where the rightmost inequality follows by definition. But (1-23) is true for any β, thus

$$S_1 = \min_{\beta \epsilon B} \max_{\alpha \epsilon A} J_M(\alpha, \beta) \geq J(\alpha, \beta^*); \;\; \forall\, \alpha \epsilon A \tag{1-24}$$

Thus the S_1 criterion places a least upper bound on the filter estimation error in the presence of uncertain parameters and may be considered a "worst case" design.

The S_2 and S_3 criteria seek to control the filter sensitivity directly by minimizing the maximum deviation of the estimation error variance from the optimum error variance over the set of uncertain parameters in either an absolute or relative sense. The value of S_2 or S_3 can be viewed as a measure of the degree to which the filter (1-13) tracks optimality.

Minimax design criteria are not new. The validity of their application to the control of plants with large parameter uncertainties has been established [8,9,10]. The extent of their applicability to the filtering problem is the major subject of this work. It must be emphasized that the design criteria proposed herein allow for "large" uncertainties in system parameters.

In what follows the S_1 filter will often be referred to as the minimax filter. The S_2 and S_3 filters will be called minimax sensitivity filters.

1.3 Thesis Organization

The organization of the remainder of this thesis is as follows: In Chapter II the properties of the minimax filter for uncertain but constant noise covariances are developed. The solution of the infinite time problem is discussed in detail and several simple examples are presented.

Chapter III sets forth the properties of minimax sensitivity filters for uncertain but constant noise covariances, again concentrating on the infinite time solution.

Chapter IV discusses computational procedures for finding S_1 filters for higher order systems and presents several illustrative examples.

Chapter V contains some partial results concerning systems with uncertain but constant dynamic parameters and systems with time varying noise statistics.

Finally, Chapter VI presents a summary of all results and suggests areas for further research.

CHAPTER II

THE MINIMAX FILTER FOR UNCERTAIN NOISE STATISTICS

2.1 Introduction

In this chapter the properties of the minimax (S_1) filter for time invariant plants with constant but uncertain input and measurement noise covariance matrices are developed.

For this problem the plant and measurement equations take the form

$$\dot{x}(t) = F\,x(t) + G\,u(t) \tag{2-1}$$

$$z(t) = H\,x(t) + w(t) \tag{2-2}$$

where

$$E\left\{u(t)\right\} = 0 \qquad E\left\{w(t)\right\} = 0 \tag{2-3}$$

$$\mathrm{Cov}\,[u(t)] = Q\,\delta(t-\tau); \quad Q \geq 0 \tag{2-4}$$

and

$$\mathrm{Cov}\,[w(t)] = R\,\delta(t-\tau); \quad R > 0 \tag{2-5}$$

The inequalities of eqs. (2-4) and (2-5) mean that R and Q are respectively positive definite and positive semi-definite matrices.

In accordance with the discussion of Chapter I the form of the minimax filter is specified as:

$$\dot{\bar{x}}(t) = F\bar{x}(t) + K(t)[z(t) - H\bar{x}(t)]; \qquad \bar{x}(t_o) = 0 \tag{2-6}$$

The filter (2-6) is easily shown to be an unbiased estimator for x(t) as follows. Substituting (2-2) into (2-6) one obtains

$$\dot{\bar{x}}(t) = [F - K(t)H]\bar{x}(t) + K(t)H\,x(t) + K(t)\,w(t) \tag{2-7}$$

Recognizing that the expected value of x(t) and w(t) are both zero the expected value of $\dot{\bar{x}}(t)$ becomes

$$E\{\dot{\bar{x}}(t)\} = \frac{d}{dt}E\{\bar{x}(t)\} = [F - K(t)H]E\{\bar{x}(t)\} \tag{2-8}$$

The solution for eq. (2-8) can be written as

$$E\{\bar{x}(t)\} = \Phi_K(t, t_o)\,E\{\bar{x}(t_o)\} \tag{2-9}$$

where $\Phi_K(t, t_o)$, the state transition matrix for eq. (2-8), satisfies the matrix differential equation

$$\dot{\Phi}_K(t, t_o) = [F - K(t)H]\Phi_K(t, t_o); \qquad \Phi_K(t_o, t_o) = I \tag{2-10}$$

From (2-6), however, $\bar{x}(t_o) = 0$. Thus $E\{\bar{x}(t_o)\} = 0$ and finally using (2-9)

$$E\{\bar{x}(t)\} = E\{x(t)\} = 0 \tag{2-11}$$

Defining the estimation error e(t) as $x(t) - \bar{x}(t)$ one has from (2-1) and (2-6)

$$\dot{e}(t) = [F - K(t)H]\,e(t) + G\,u(t) - K(t)\,w(t) \tag{2-12}$$

Since the filter (2-6) is an unbiased estimator for x(t) one has immediately

$$E\{e(t)\} = 0 \tag{2-13}$$

Let

$$M(t) = \mathrm{Cov}[e(t)] = E\{e(t)\,e^T(t)\} \tag{2-14}$$

Then M(t) satisfies the linear matrix differential equation

$$\dot{M}(t) = [F - K(t)H]\,M(t) + M(t)[F - K(t)H]^T + GQG^T$$
$$+ K(t)\,RK^T(t); \qquad M(t_o) = P_o \tag{2-15}$$

which has as its solution

$$M(t) = \Phi_K(t,t_o)\,P_o\Phi_K^T(t,t_o) + \int_{t_o}^{t} \Phi_K(t,\sigma)\left[GQG^T + K(\sigma)RK^T(\sigma)\right]\Phi_K(t,\sigma)\,d\sigma \tag{2-16}$$

The filter performance measure first given in Chapter I is now generalized somewhat to include an arbitrary positive definite weighting matrix W.[1]

---

[1] This restriction guarantees the uniqueness of the Kalman filter [see eq. (5-15)]. In many instances, however, $W \geq 0$ is sufficient (see examples 4-2 and 4-3 of Chapter IV).

$$J_M(T) = \mathrm{tr}\,[WM(T)] = \mathrm{tr}\,[WM(K, Q, R, P_o, t_o, T)] \tag{2-17}$$

K, Q, R, and P_o are considered independent variables. Eqs. (2-15), (2-16) and (2-17) provide a complete mathematical description of the error performance of filter (2-6).

2.2 Problem Definition

Before defining the minimax filter for uncertain noise statistics it is necessary to discuss the nature of the uncertainties to be allowed in the R and Q matrices. Of course R and Q must be positive definite and semi-definite respectively. Beyond this, however, in order to derive maximum benefit from the minimax formulation of the filtering problem, it is desirable to place as few additional restrictions as possible upon the allowed class of R and Q matrices consistent with the need of obtain a well-posed problem. Ideally these restrictions should reflect physical limitations of the problem at hand.

The diagonal terms of Q and R represent in some sense the mean square power in each component of the plant and measurement noise vectors. Using either physical considerations or limited experimental data it is not too difficult to place reasonable upper and lower bounds on the power in each component. This in turn establishes upper and lower bounds on the trace of Q and R. This single restriction on the trace of Q and R will be sufficient for our purposes. Accordingly, the following sets are defined:

$$V_Q = \left\{ Q \mid Q = Q^T, \ Q \geq 0, \text{ and } 0 \leq q_{min} \leq \operatorname{tr} Q \leq q_{max} < \infty \right\} \tag{2-18}$$

and

$$V_R = \left\{ R \mid R = R^T, \ R > 0, \text{ and } 0 < r_{min} \leq \operatorname{tr} R \leq r_{max} < \infty \right\} \tag{2-19}$$

Thus V_Q is the set of all real symmetric positive semi-definite matrices with finite trace lying between q_{min} and q_{max} and V_R is the set of all real symmetric positive definite matrices with finite trace lying between r_{min} and r_{max}. For notational convenience the set $V = V_Q \times V_R$ is defined and elements of this set are denoted by v.

The matrix K(t) of eq. (2-6) is an independent variable to be selected by the designer. For our purposes K(t) may be in any compact convex set $V_K \in E_{np}$ which covers the set $\mathcal{K}_o$ defined as

$$\mathcal{K}_o = \left\{ K \mid K = K_o(t) \text{ for some } v \in V \text{ and } t_o \leq t \leq T \right\} \tag{2-20}$$

This requirement insures that V_K will be large enough to generate the Kalman filter for any admissible value of Q and R. To show that a set $V_K \supset \mathcal{K}_o$ exists it is sufficient to show that $\mathcal{K}_o$ is bounded. Using the matrix norm defined in Appendix I and eq. (1-7)

$$\|K_o\|^2 = \mathrm{tr}(PH^TR^{-2}HP) = \mathrm{tr}(HP^2H^TR^{-2})$$
$$\leq \|HP^2H^T\| \cdot \|R^{-2}\|$$
$$\leq \|H^TH\| \cdot \|P\|^2 \cdot \|R^{-1}\|^2$$

or

$$\|K_o\| \leq \frac{1}{r_{min}} \|H^TH\|^{\frac{1}{2}} \cdot \|P\| \tag{2-21}$$

P(t) is known to exist for all $t \geq t_o$ and every $v \in V$ and any non-negative definite P_o. Thus $\|P\|$ is bounded and by (2-21) so is K_o.

The minimax filter for uncertain but constant Q and R may now be defined.

Def. 2.1: The minimax filter for uncertain constant Q and R is defined by that element of V_K denoted $K^*(T)$ for which

$$\max_{v \in V} J_M(K^*(T), v, P_o, t_o, T) = \min_{K \in V_K} \max_{v \in V} J_M(K, v, P_o, t_o, T) \tag{2-22}$$

The major result of this chapter will be to show that a saddle point exists for this problem. That is

$$\min_{K \in V_K} \max_{v \in V} J_M(K, v, P_o, t_o, T) = \max_{v \in V} \min_{K \in V_K} J_M(K, v, P_o, t_o, T) \tag{2-23}$$

Observe that the right-hand side of (2-23) can be written as

$$\max_{v \epsilon V} \left[\min_{K \epsilon V_K} J_M(K, v, P_o, t_o, T) \right] = \max_{v \epsilon V} J_o(v, P_o, t_o, T) \tag{2-24}$$

Thus the relatively difficult problem of minimaximizing J_M can be replaced with the simpler problem of maximizing J_o over the uncertain parameter set V.

2.3 Convexity and Game Theory

As a preliminary to demonstrating (2-23) some properties of convex sets and convex functions are first reviewed and two results from game theory are cited. Most of this material is available in the references [11, 12]. Where lacking, however, short proofs of some properties are presented.

Def. 2.2: A set of points X is said to be convex if whenever two points x_1 and x_2 belong to X, all points of the form $\alpha x_1 + (1-\alpha)x_2$, $0 \le \alpha \le 1$, also belong to X.

Def. 2.3: A scalar function f(x) defined on a convex subset X of E_n is said to be convex in X if

$$f[\alpha x_1 + (1-\alpha)x_2] \le \alpha f(x_1) + (1-\alpha) f(x_2)$$

for every x_1 and x_2 in X and any scalar α, $0 \le \quad \le 1$. If

$$f[\alpha x_1 + (1-\alpha)x_2] < \alpha f(x_1) + (1-\alpha) f(x_2)$$

then f(x) is said to be strictly convex. f(x) is concave if -f(x) is convex. If f(x) is linear in X it is both concave and convex.

Property 2.1: If f(x) is a twice differentiable function and X is a convex set in E_n, then f(x) is convex if and only if the matrix of second derivatives

$$f_{xx} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{i,j}$$

is positive semi-definite for every $x \epsilon X$. If the above matrix is positive definite, then f(x) is strictly convex.

Lemma 2.1: If f(x, y, t) is a continuous function of x, y, and t and if for every y and t, f(x, y, t) is a convex function of $x \epsilon X$ with second order partials with respect to x continuous in y and t then

$$g(x, y) = \int_{t_o}^{T} f(x, y, t)\, dt$$

is a convex function of x for every y.

Proof: Under the conditions of the lemma one may differentiate under the integral sign to obtain

$$g_{xx} = \frac{\partial^2 g(x, y)}{\partial x^2} = \int_{t_o}^{T} f_{xx}(x, y, t)\, dt \tag{2-25}$$

Consider the associated quadratic form

$$\left\langle z, g_{xx} z \right\rangle = \int_{t_o}^{T} \left\langle z, f_{xx} z \right\rangle dt \tag{2-26}$$

Since the integrand is non-negative for every $z \in E_n$ it follows that $\left\langle z, g_{xx} z \right\rangle \geq 0$ and therefore $g_{xx} \geq 0$. By property ~~1,~~ g(x, y) is convex in X for every y. If f(x, y, t) is strictly convex in X then so is g(x, y).

Two well-known theorems from game theory are now cited [12].

Theorem 2.1: Let f(x, y) be a real-valued function of two variables x and y which are elements of X and Y respectively, where both X and Y are closed, bounded, convex sets. If f is continuous, convex in y for each x and concave in x for each y, then

$$\min_{y \in Y} \max_{x \in X} f(x, y) = \max_{x \in X} \min_{y \in Y} f(x, y)$$

Theorem 2.2: A necessary and sufficient condition that

$$\min_{y \in Y} \max_{x \in X} f(x, y) = \max_{x \in X} \min_{y \in Y} f(x, y)$$

is that there exist unique points x_o and y_o such that for all $y \in Y$ and all $x \in X$

$$f(x, y_o) \leq f(x_o, y_o) \leq f(x_o, y)$$

Theorem 2.2 is often referred to as the saddle point theorem and (x_o, y_o) is called the saddle point of $f(x, y)$.

2.4 The Minimax Filter

The minimax filter problem formulated in Section 2 will now be shown to satisfy all the sufficient conditions of theorem 2.1.

The set V_K is convex, bounded, and closed by definition. The convexity of V_Q is easily demonstrated as follows: Given $Q_1, Q_2 \in V_Q$ then for $0 \leq \alpha \leq 1$

$$\alpha q_{min} \leq \alpha \operatorname{tr} Q_1 \leq \alpha q_{max} \tag{2-27a}$$

and

$$(1 - \alpha) q_{min} \leq (1 - \alpha) \operatorname{tr} Q_1 \leq (1 - \alpha) q_{max} \tag{2-27b}$$

Adding inequalities (2-27) and using the linearity of the trace operator yields

$$q_{min} \leq \operatorname{tr} [\alpha Q_1 + (1 - \alpha) Q_2] \leq q_{max} \tag{2-28}$$

Now since a positive scalar multiple of a positive semi-definite matrix is positive semi-definite and since the sum of two positive semi-definite matrices is also positive semi-definite

$$Q = \alpha Q_1 + (1-\alpha)Q_2 \epsilon V_Q \tag{2-29}$$

and V_Q is convex.

An identical development demonstrates the convexity of V_R. Finally, since the Cartesian product of two convex sets is convex, V is a convex set. V is shown to be bounded and closed in Appendix A.

Continuity of J_M in V_K and V for all finite t is easily demonstrated by a straightforward though tedious application of the properties of the matrix norms defined in Appendix A. Since J_M in linear in V_Q and V_R it is by definition 2.3 concave in V. The reader should observe that J_M is continuous and concave in P_o also so that the present development is valid for uncertainties in initial covariance merely be redefining V to include such uncertainties.

It remains to be shown that J_M is convex in V_K for every $v \epsilon V$ and $t \geq t_o$. This is done by showing that the time derivative of J_M is convex in V_K and then invoking lemma 2.1. From eq. (2-15) one has

$$\dot{J}_M = \mathrm{tr}\,(W\dot{M}) = \mathrm{tr}\left[W(F-KH)M + WM(F-KH)^T + WGQG^T + WKRK^T\right] \tag{2-30}$$

Using the concept of the gradient matrix [13] the first partial of $\dot{J}_M$ with respect to K is

$$\frac{\partial \dot{J}_M}{\partial K} = -2WMH^T + 2WKR \tag{2-31}$$

In order to avoid tensor notation in the second differentiation the matrix K will be mapped into an equivalent column vector as follows:[2]

Let

$$K = \begin{bmatrix} k_1^T \\ k_2^T \\ \vdots \\ k_m^T \end{bmatrix} \tag{2-32}$$

where

$$k_i^T = (k_{i1}, k_{i2}, \dots k_{ip}) \tag{2-33}$$

and denote by k the np x 1 column vector

$$k = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_m \end{bmatrix} \tag{2-34}$$

[2] The reader should recognize that what follows is merely notational manipulation since K is actually a vector in E_{np}.

Then eq. (2-31) becomes

$$\frac{\partial \dot{J}_M}{\partial k} = -\gamma + 2(W \otimes R)\,k \tag{2-35}$$

where γ is the appropriate matrix to vector mapping of $2WMH^T$ and $\otimes$ denotes Bellman's Kronecker product for matrices [14]. $W \otimes R$ is the $np \times np$ matrix

$$W \otimes R = \left(w_{ij}R\right) \tag{2-36}$$

The second partial of $\dot{J}_M$ with respect to K is now simply

$$\frac{\partial^2 \dot{J}_M}{\partial k^2} = 2(W \otimes R) \tag{2-37}$$

The eigenvalues of $W \otimes R$ are

$$\lambda_{ij}(W \otimes R) = \lambda_i(W)\,\lambda_j(R); \quad 1 \le i \le n, \ \ 1 \le j \le p \tag{2-38}$$

But $\lambda_i(W) > 0$ and $\lambda_j(R) > 0$ so that $\lambda_{ij}(W \otimes R) > 0$ and $W \otimes R$ is positive definite. Thus $\dot{J}_M$ is strictly convex in V_K. Now

$$M(T) = P_o + \int_{t_o}^{T} \dot{M}(t)\,dt \tag{2-39}$$

so that

$$J_M(T) = \mathrm{tr}\,(WP_o) + \mathrm{tr}\left[\int_{t_o}^{T} W\dot{M}(t)\,dt\right]$$

$$= \mathrm{tr}\,(WP_o) + \int_{t_o}^{T} \dot{J}_M\,dt \qquad (2\text{-}40)$$

The convexity of the integral of $\dot{J}_M$ follows immediately from lemma 2.1. Since the addition of a constant to any convex function does not affect its convexity, the convexity of J_M in V_K is established.

Thus all the conditions of theorem 2.1 are met and the assertion (2-23) is true, namely that

$$\min_{K\epsilon V_K}\ \max_{v\epsilon V}\ J_M(K, v, P_o, t_o, T) = \max_{v\epsilon V}\ \min_{K\epsilon V_K}\ J_M(K, v, P_o, t_o, T)$$

Using Def. 2.1 and eqs. (1-20) and (2-24) one obtains

$$S_1(P_o, t_o, T) = \max_{v\epsilon V}\ J_M(K^*(T), v, P_o, t_o, T) = \max_{v\epsilon V}\ J_o(v, P_o, t_o, T)$$

or

$$S_1(P_o\ t_o, T) = J_o(v_T^*, P_o, t_o, T) \qquad (2\text{-}41)$$

where v_T^* is the maximizing v at time T.

Theorem 2.2 and eq. (2-41) together imply that

$$K^*(t) = K_o(v_T^*, t); \qquad t_o \le t \le T \qquad (2\text{-}42)$$

2.5 The Infinite Time Minimax Filter

Although the minimax problem has been replaced with a simpler maximization over the set of all Kalman filter responses, the problem for arbitrary T is still a formidable one. The dependence of J_o on P_o further complicates matters since any uncertainty in P_o would necessitate its inclusion in the set of uncertain parameters. Under fairly general conditions (see theorem 2.3 below) J_o is independent of P_o as T becomes arbitrarily large. By limiting our attention to the infinite time case it should be possible to remove the P_o dependence from our problem. A further advantage of the infinite time case is that the equation for P_o is algebraic rather than differential and the resulting algebraic maximum is relatively easy to find.

The following important result due to Kalman is now cited [2]:

Theorem 2.3: Assume the following is true of system (2-1) and (2-2):

(a) uniformly completely observable

(b) uniformly completely controllable

(c) F, G, H, Q, and R are constant bounded matrices.

Then:

(i) Every solution of the variance equation (1-9) starting at a symmetric non-negative matrix P_o converges to a unique constant non-negative matrix $\overline{P}$ as $t \to \infty$.

(ii) The optimal filter is uniformly asymptotically stable.

For time invariant plants $P(v, P_o, t_o, T)$ and therefore $J_o(v, P_o, t_o, T)$ are functions of $T - t_o$ only. For convenience $t_o = 0$ will be used from now on.

Let us now examine eq. (2-41) in the limit as T becomes arbitrarily large. Taking the limit of both sides as $T \to \infty$ one obtains

$$\lim_{T\to\infty} S_1(P_o, T) = \lim_{T\to\infty} J_o(v_T^*, P_o, T) \tag{2-43}$$

The proof of theorem 2.3 requires only that Q and R be bounded in norm. The set V is bounded in norm. Furthermore, since V is compact the convergence of P(T) to $\overline{P}$ and therefore $J_o(T)$ to $\overline{J}_o$ is uniform in V. That is, there exists a $T(\epsilon)$ such that for every $T > T(\epsilon)$ and any $v \in V$

$$|J_o(v, P_o, T) - \overline{J}_o(v)| < \epsilon \tag{2-44}$$

Now J_o is always non-negative so eq. (2-44) implies that

$$-\epsilon + J_o(v, P_o, T) < \overline{J}_o(v) \le J_o(v_\infty^*) \tag{2-45}$$

where the rightmost inequality in (2-45) follows from the definition of v_T^*. Since (2-45) is true for every $v \in V$ one has

$$-\epsilon + J_o(v_T^*, P_o, T) < \overline{J}_o(v_\infty^*)$$

or

$$J_0(v_T^*, P_0, T) - \bar{J}_0(v_\infty^*) < \epsilon \tag{2-46}$$

Again from eq. (2-44) and the definition of v_T^*

$$\bar{J}_0(v_\infty^*) < \epsilon + J_0(v_\infty^*, P_0, T) \leq \epsilon + J_0(v_T^*, P_0, T) \tag{2-47}$$

which is equivalent to

$$\bar{J}_0(v_\infty^*) - J_0(v_T^*, P_0, T) < \epsilon \tag{2-48}$$

Eqs. (2-46) and (2-48) together imply that

$$|J_0(v_T^*, P_0, T) - \bar{J}_0(v_\infty^*)| < \epsilon; \quad T > T(\epsilon) \tag{2-49}$$

Eq. (2-49) is equivalent to the statement

$$\lim_{T\to\infty} J_0(v_T^*, P_0, T) = \bar{J}_0(v_\infty^*)$$

or

$$\lim_{T\to\infty} \max_{v\epsilon V} J_0(v, P_0, T) = \max_{v\epsilon V} \lim_{T\to\infty} J_0(v, P_0, T) \tag{2-50}$$

Thus the operations of "max" and "lim" may be interchanged and eq. (2-43) becomes

$$\lim_{T\to\infty} S_1(P_0, T) \equiv S_1(\infty) = \max_{v\epsilon V} \bar{J}_0(v) \tag{2-51}$$

or

$$\lim_{T\to\infty} \min_{K\epsilon V_K} \max_{v\epsilon V} J_M(K, v, P_0, T) = \max_{v\epsilon V} \bar{J}_0(v) \tag{2-52}$$

Thus the minimax value of J_M as T becomes arbitrarily large is equal to the maximum over V of the infinite time optimal return and from theorem 2.3

$$\lim_{T\to\infty} K^*(T) = K_\infty^* = \overline{K}_0(v_\infty^*) \tag{2-53}$$

In the general formulation of the minimax filtering problem given in Chapter I the filter matrix $F(\beta_1)$ was allowed to differ from the system matrix $F(\alpha_1)$. Since $F(\alpha_1)$ was assumed known in this chapter, $F(\beta_1)$ was set equal to $F(\alpha_1)$. The question naturally arises as to whether or not a smaller value for S_1 can be attained by allowing $F(\beta_1)$ to vary. This question is easily answered.

Let $F(\beta_1)$ be restricted to a compact set V_F containing F and denote by c, the elements of the set

$$C = V_F \times V_K \tag{2-54}$$

Let $J_M'(c, v, P_o, T)$ denote the performance index for this new filter. Now it is always true that [12]:

$$\min_{c\epsilon C}\ \max_{v\epsilon V}\ J_M'(c, v, P_o, T) \geq \max_{v\epsilon V}\ \min_{c\epsilon C}\ J_M'(c, v, P_o, T) \tag{2-55}$$

But $\min_{c\epsilon C} J_M'(c, v, P_o, T)$ clearly occurs when $c = [F(\alpha_1), K_o]$ and has the value $J_o(v, P_o, T)$ so that (2-55) becomes

$$\min_{c \epsilon C} \max_{v \epsilon V} J'_M(c, v, P_o, T) = \max_{v \epsilon V} J_o(v, P_o, T) \tag{2-56}$$

Comparing (2-56) with (2-41) it is clear that no smaller minimax is obtained by allowing the filter and system F matrices to differ.

2.6 A Maximization Problem

We have seen that the minimax problem may be replaced with a simple maximization of J_o over the unknown parameter set. In this section some of the general properties of this maximization for the infinite time case are discussed.

The steady-state covariance of the optimum estimate for some nominal Q_N and R_N is obtained by equating the left-hand side of eq. (1-9) to zero [2]. Thus

$$0 = F\overline{P}_N + \overline{P}_N F^T + GQG^T - \overline{P}_N H^T R_N^{-1} H\overline{P}_N \tag{2-57}$$

An expression for the gradient of tr $[W\overline{P}_N]$ with respect to an uncertain element of the Q or R matrix will not be developed. Consider first an uncertain element in Q, say q_{ij}. Let

$$Q_{ij} = \frac{\partial Q}{\partial q_{ij}} \tag{2-58a}$$

and

$$\overline{P}_{ij} = \frac{\partial \overline{P}_N}{\partial q_{ij}} \tag{2-58b}$$

Since Q is symmetric, Q_{ij} is symmetric with two forms:

$$Q_{ii} = \left\{1_{ii}\right\} \quad i = j \tag{2-59a}$$

or

$$Q_{ij} = \left\{1_{ij}\right\} + \left\{1_{ji}\right\} \quad i \neq j \tag{2-59b}$$

where $\left\{1_{ij}\right\}$ denotes a matrix whose entries are all zero except for a 1 in the ij<u>th</u> position.

The total derivative of (2-57) with respect to q_{ij} is then

$$0 = F\overline{P}_{ij} + \overline{P}_{ij}F + GQ_{ij}G^T - \overline{P}_{ij}H^TR_N^{-1}H\overline{P}_N - \overline{P}_NH^TR_N^{-1}H\overline{P}_{ij} \tag{2-60}$$

Regrouping (2-60) yields

$$0 = \left(F - \overline{P}_NH^TR_N^{-1}H\right)\overline{P}_{ij} + \overline{P}_{ij}\left(F - P_NH^TR_N^{-1}H\right)^T + GQ_{ij}G^T \tag{2-61}$$

Observe that (2-61) is a linear matrix algebraic equation for $\overline{P}_{ij}$. This equation has a solution whenever the eigenvalues of $(F - P_NH^TR_N^{-1}H)$ are non-zero [14]. But this matrix is the system matrix of the steady-state Kalman filter which by theorem 2.3 is uniformly asymptotically stable. Thus the eigenvalues of $(F - \overline{P}_NH^TR_N^{-1}H)$ are non-zero and $\overline{P}_{ij}$ always exists.

The solution to (2-61) is [14]:

$$\overline{P}_{ij} = \int_0^\infty \Phi_K(\sigma)\, GQ_{ij}G^T\Phi_K(\sigma)\, d\sigma \tag{2-62}$$

where

$$\Phi_K(\sigma) = e^{\left(F - \overline{P}_N H^T R_K^{-1} H\right)\sigma} \tag{2-63}$$

Now Q_{ij} is positive semi-definite when $i = j$. In this case the integrand in (2-62) is positive semi-definite for every σ and thus $\overline{P}_{ij}$ is positive semi-definite. This last fact implies that

$$\left\{ \overline{P}_{ii} \right\}_{\ell\ell} \geq 0; \quad 1 \leq \ell \leq n; \quad 1 \leq i \leq m \tag{2-64}$$

and

$$\frac{\partial \overline{J}_o}{\partial q_{ii}} = \mathrm{tr}\,(W\overline{P}_{ii}) > 0 \tag{2-65}$$

The above results may be summarized as follows:

(1) $\overline{P}_{ij}$ and therefore $\partial \overline{J}_o / \partial q_{ij}$ always exist.

(2) $\partial \overline{J}_o / \partial q_{ii}$ is always positive and thus in maximizing $\overline{J}_o$ all diagonal elements of Q take on their maximum allowed value.[3]

(3) Maximizing $\overline{J}_o$ with respect to a diagonal term of Q also maximizes each element of tr $[\overline{P}]$ with respect to that term in Q (let W = I).

---

[3] For $W \geq 0$, $\overline{J}_o$ may be independent of one or more diagonal elements of Q.

$\overline{P}_{ii}$ can be interpreted as the steady-state covariance of the state of the infinite time Kalman filter for Q_N and R_N when excited at the ith input by a white noise of unit variance all other inputs being zero. Thus tr $(W\overline{P}_{ii})$ is the change in tr $(W\overline{P}_N)$ per unit variance change in u_i.

Unfortunately when $i \neq j$, Q_{ij} is indefinite and therefore so is $\overline{P}_{ij}$. Thus no general statement concerning the maximizing value of q_{ij} can be made.

Let us now turn our attention to uncertain elements in the matrix R. Let r_{ij} denote the elements of R and define

$$R_{ij} = \frac{\partial R}{\partial r_{ij}} \tag{2-66}$$

where R_{ij} is defined similarly to Q_{ij} in (2-59). Also note that

$$\frac{\partial R^{-1}}{\partial r_{ij}} = -R^{-1} R_{ij} R^{-1} \tag{2-67}$$

The total derivative of (2-57) with respect to r_{ij} is then

$$\begin{aligned} 0 = F\overline{P}_{ij} + \overline{P}_{ij}F^T + \overline{P}_N H^T R_N^{-1} R_{ij} R_N^{-1} H \overline{P}_N \\ - \overline{P}_{ij} H^T R_N^{-1} H \overline{P}_N - \overline{P}_N H^T R_N^{-1} H \overline{P}_{ij} \end{aligned} \tag{2-68}$$

where now

$$\overline{P}_{ij} = \frac{\partial \overline{P}_N}{\partial r_{ij}} \tag{2-69}$$

Regrouping (2-68) one obtains

$$0 = \left(F - \overline{P}_N H^T R_N^{-1} H\right)\overline{P}_{ij} + \overline{P}_{ij}\left(F - \overline{P}_N H^T R_N^{-1} H\right)^T$$
$$+ \overline{P}_N H^T R_N^{-1} R_{ij} R_N^{-1} H \overline{P}_N \qquad (2\text{-}70)$$

Finally, since $\overline{K}_N = \overline{P}_N H^T R_N^{-1}$, (2-70) becomes

$$0 = (F - \overline{K}_N H)\overline{P}_{ij} + \overline{P}_{ij}(F - \overline{K}_N H) + \overline{K}_N R_{ij} \overline{K}_N^T \qquad (2\text{-}71)$$

As in the case for uncertain Q, (2-71) is a linear matrix algebraic equation for $\overline{P}_{ij}$ involving the steady state Kalman filter system matrix $(F - \overline{K}_N H)$. Since this matrix has no zero eigenvalues (2-71) always has a solution. This solution is

$$\overline{P}_{ij} = \int_0^{\infty} \Phi_K(\sigma)\overline{K}_N R_{ij} \overline{K}_N^T \Phi^T(\sigma)\, d\sigma \qquad (2\text{-}72)$$

where $\Phi_K(\sigma)$ is defined in (2-63).

When $i = j$, R_{ij} is positive semi-definite and therefore so is $\overline{P}_{ii}$. $\overline{P}_{ii}$ can be viewed as the steady-state covariance of the state of the infinite time Kalman filter in response to a measurement noise w_i of unit variance, all other inputs being zero. Thus tr $[W\overline{P}_{ii}]$ is the change in tr $[W\overline{P}_N]$ per unit variance change in w_i. It is clear then that

$$\left\{\overline{P}_{ii}\right\}_{\ell\ell} \geq 0; \quad 1 \leq \ell \leq n, \quad 1 \leq i \leq p \qquad (2\text{-}73)$$

and

$$\frac{\partial \bar{J}_o}{\partial r_{ii}} = \operatorname{tr}(W\bar{P}_{ii}) > 0 \tag{2-74}$$

The preceding results are summarized as follows:

(1) $\bar{P}_{ij}$ and therefore $\partial\bar{J}_o/\partial r_{ij}$ always exist.

(2) $\partial\bar{J}_o/\partial r_{ii}$ is always positive and thus in maximizing $\bar{J}_o$ all diagonal terms of R take on their maximum value.

(3) Maximization of $\bar{J}_o$ with respect to a diagonal term in R also maximizes each diagonal term of $\bar{P}$ with respect to that term in R.

When $i \neq j$, R_{ij} is indefinite and therefore so is $\bar{P}_{ij}$. Thus no general conclusion concerning the maximizing value of r_{ij} can be made.

In the derivations of this section it has tacitly been assumed that the various elements of V are independent. These elements are, however, constrained by the requirement that Q and R be respectively positive semi-definite and positive definite. These constraints are embodied in the well-known inequalities concerning the principal minors of Q and R [15]. Since these inequalities are highly nonlinear, determining the region in V_Q or V_R for which they are satisfied is an extremely difficult problem. Furthermore, since the constraints are fewer in number than the off-diagonal terms, the bounds one obtains are not absolute, but merely nonlinear functional relationships between terms.

A simple technique is presented in Chapter III whereby a range for each off-diagonal term is determined such that within that range a matrix with fixed diagonal elements is always positive definite or semi-definite as required. Thus in this range the off-diagonal terms are independent. Although this technique limits somewhat the maximum degree of cross-correlation one may consider between any two components of u or w, it greatly simplifies computation of the maximum.

2.7 Some Examples

At this point a few simple examples illustrating the theory so far developed are in order.

Example 2.1

Consider the simple first-order plant

$$\dot{x} = -x + u \tag{2-75}$$

with noisy measurement

$$z = x + w \tag{2-76}$$

where

$$E\{u\} = E\{w\} = 0$$

and

$$\mathrm{Cov}\,[u] = q\delta(t-\tau) \tag{2-77}$$

$$\mathrm{Cov}\,[w] = r\delta(t-\tau) \tag{2-78}$$

The infinite time minimax filter takes the form

$$\dot{\bar{x}} = -\bar{x} + k(z - \bar{x}) \qquad (2\text{-}79)$$

From eq. (2-15) the derivative of the estimation error covariance is

$$\dot{m} = -2(1 + k)m + q + k^2 r \qquad (2\text{-}80)$$

Assuming $k > -1$, the steady-state value of m is found by setting the left-hand side of (2-80) equal to zero, yielding

$$m = \frac{k^2 r + q}{2(1 + k)} \qquad (2\text{-}81)$$

Observe that m is linear in r and q and thus the maximizing values of r and q are r_{max} and q_{max} respectively. Also

$$\frac{\partial^2 m}{\partial k^2} = \frac{r + q}{(1 + k)^3} > 0 \qquad (2\text{-}82)$$

for $r > 0$ and $q > 0$, so that m is strictly convex in k. Now the minimizing value of k is found from

$$\left.\frac{\partial m}{\partial k}\right|_{\substack{r = r_{max} \\ q = q_{max}}} = \frac{k^2 r_{max} + 2kr_{max} - q_{max}}{2(1 + k)^2} = 0 \qquad (2\text{-}83)$$

which yields

$$k^* = -1 \pm \left(1 + \frac{q_{max}}{r_{max}}\right)^{\frac{1}{2}} \tag{2-84}$$

In order to maintain the stability of the filter (2-79) for arbitrary r and q one must choose the "+" sign in (2-84). Substituting (2-84) back into (2-81), one obtains

$$m^* = \left(r_{max}^2 + r_{max}q_{max}\right)^{\frac{1}{2}} - r_{max} \tag{2-85}$$

For known r and q the optimal covariance is obtained from eq. (1-9).

$$\dot{p} = -2p + q - \frac{p^2}{r} \tag{2-86}$$

Equating (2-86) to zero yields

$$\bar{p} = (r^2 + rq)^{\frac{1}{2}} - r \tag{2-87}$$

Now

$$\frac{\partial \bar{p}}{\partial q} = \frac{1}{2} \frac{r}{(r^2 + rq)^{\frac{1}{2}}} > 0 \tag{2-88}$$

and

$$\frac{\partial \bar{p}}{\partial r} = \frac{2r + q}{2(r^2 + rq)^{\frac{1}{2}}} - 1 > 0 \tag{2-89}$$

The last inequality is seen most directly be squaring the expression

$$\frac{2r + q}{2(r^2 + rq)^{\frac{1}{2}}} \tag{2-90}$$

which produces

$$\frac{4r^2 + 4rq + q^2}{4r^2 + 4rq} = 1 + \frac{q^2}{4(r^2 + rq)} > 1 \tag{2-91}$$

From (2-88) and (2-89) it is seen that

$$\max_{r,q} \bar{p}(r, q) = (r_{max}^2 + r_{max}q_{max})^{\frac{1}{2}} - r_{max} \tag{2-92}$$

Furthermore the Kalman gain associated with this value of $\bar{p}$ is

$$k_o = \frac{\bar{p}(r_{max}, q_{max})}{r_{max}} = \left(1 + \frac{q_{max}}{r_{max}}\right)^{\frac{1}{2}} - 1 \tag{2-93}$$

Thus minmax has been shown equal to maxmin for this simple example by direct calculation and from (2-84) and (2-93) it is clear that

$$k^* = k_o(r_{max}, q_{max}) \tag{2-94}$$

Let us now examine the error surfaces of the optimal and minimax filters over the $(q \times r)$ space. From equation (2-87) the following are apparent:

$$\bar{p}(r, q)\Big|_{r=q=0} = 0 \qquad (2\text{-}95)$$

$$\bar{p}(r, q)\Big|_{r\equiv 0} \equiv 0 \qquad (2\text{-}96)$$

$$\bar{p}(r, q)\Big|_{q\equiv 0} \equiv 0 \qquad (2\text{-}97)$$

Recall that $\bar{p}$ is concave in r and q while m is linear in these same variables. Using (2-81) one obtains:

$$m(r, q)\Big|_{r=q=0} = 0 \qquad (2\text{-}98)$$

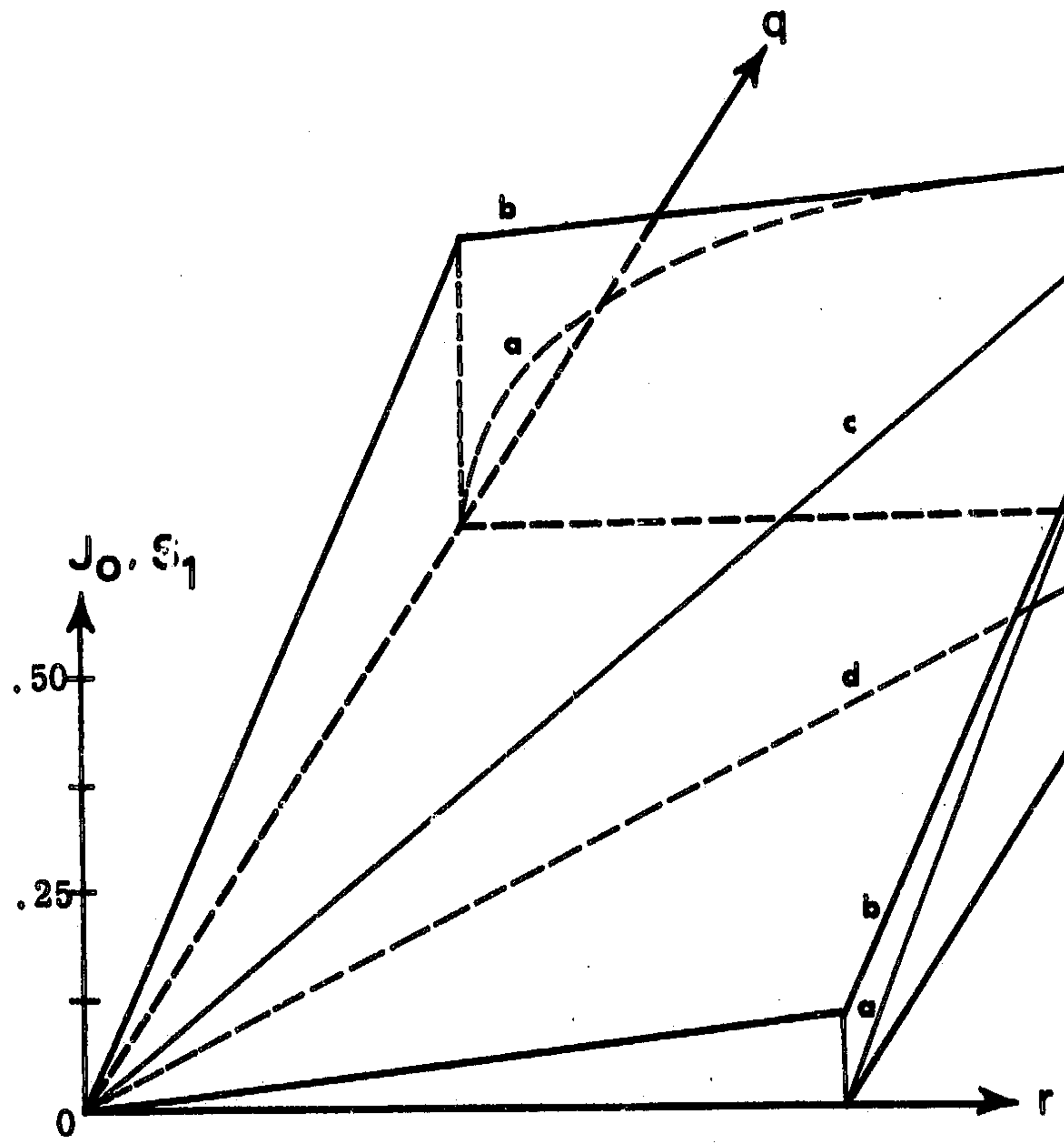
$$m(r, q)\Big|_{r=0} = \frac{q}{2(1 + k^*)} \qquad (2\text{-}99)$$

$$m(r, q)\Big|_{q=0} = \frac{(k^*)^2 r}{2(1 + k^*)} \qquad (2\text{-}100)$$

The optimal and minimax filter error surfaces are shown in Fig. 2-1. Since the Kalman gain is a function only of the ratio q/r the minimax filter is actually optimal for all values of q/r such that

$$\frac{q}{r} = \frac{q_{max}}{r_{max}} \qquad (2\text{-}101)$$

Thus the minimax error surface (a plane) has a line of contact with the



a J_o (optimal) error surface
b S_1 filter error plane
c S_1, J_o line of intersection
d projection of c on $q \times r$

Figure 2-1 S_1 and Optimal Error Surfaces for Example 2.1

optimal error surface passing through zero. The projection of this line of contact onto the $r \times q$ plane is given by eq. (2-101).

Fig. 2-1 suggests that the maximum deviation of the minimax filter error from the optimal error occurs at r_{max} or q_{max}. This is indeed true. It is merely a special result following from the convexity of the absolute sensitivity measure S^A in V. The details of this result will be developed in the next chapter. The point to remember here is that the maximum value of S^A or S^R for the minimax filter is easily found by a search over a finite number of known points in V.

Example 1.2

Here we examine the double integrator plant with scalar measurement shown in Fig. 2-2.

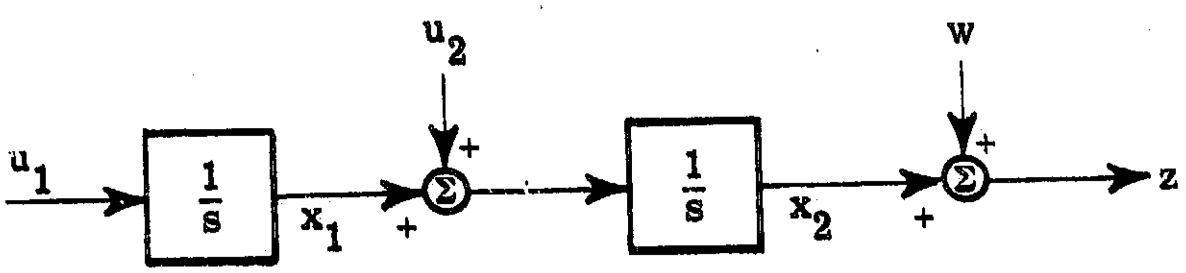


Figure 2-2 Block Diagram for Example 2.2

For this system the F, G, H, Q and R matrices are:

$$F = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}; \quad G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad H = [1 \quad 0]; \quad Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix}; \quad R = [r] \tag{2-102}$$

Using (1-9) the elements of $\overline{P}$ are found to satisfy the following set of nonlinear algebraic equations:

$$2p_{12} + q_{11} - \frac{p_{11}^2}{r} = 0 \tag{2-103}$$

$$p_{22} + q_{22} - \frac{p_{11}p_{12}}{r} = 0 \tag{2-104}$$

$$q_{22} - \frac{p_{12}^2}{r} = 0 \tag{2-105}$$

From (2-105)

$$p_{12} = \pm\sqrt{r\,q_{22}} \tag{2-106}$$

Putting (2-106) in (2-103) yields

$$p_{11}^2 = \pm 2r\sqrt{r\,q_{22}} + r\,q_{11} \tag{2-107}$$

To guarantee the positivity of p_{11}^2 for all q_{22} and q_{11} the "+" sign must be used in (2-107). Thus

$$p_{11} = \sqrt{2r\sqrt{r\,q_{22}} + r\,q_{11}} \tag{2-108}$$

and

$$p_{22} = \sqrt{q_{11}q_{22} + 2q_{22}\sqrt{r\, q_{22}}} + q_{12} \tag{2-109}$$

It is seen that the sum p_{11} and p_{22} is maximized by setting $r = r_{max}$, $q_{11} = q_{11max}$, q_{22max} and $q_{12} = q_{12max}$. Of course, q_{12max} must satisfy the constraint

$$q_{11max}q_{22max} - (q_{12max})^2 \geq 0 \tag{2-110}$$

Example 2.3

This example involves a 2×2 R matrix with an uncertain off-diagonal term. Fig. 2-3 shows a simple first-order plant with two independent measurements of the output. The plant and measurement equations are:

$$\dot{x} = -x + u \tag{2-111}$$

$$z_1 = x + w_1 \tag{2-112}$$

$$z_2 = x + w_2 \tag{2-113}$$

and the appropriate matrices are

$$F = [-1]; \quad Q = [q]; \quad H = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \quad R = \begin{bmatrix} r_{11} & r_{12} \\ r_{12} & r_{22} \end{bmatrix}$$

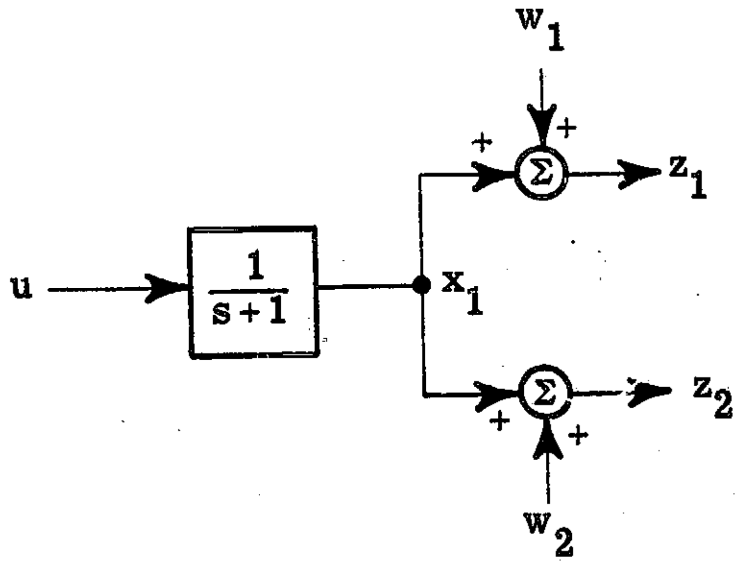


Figure 2-3 Block Diagram for Example 2.3

Putting these in the optimal covariance equation (1-9) one obtains

$$\dot{p} = -2p + q - \frac{p^2}{\left(\frac{r_{11} + r_{22} - 2r_{12}}{r_{11}r_{12} - r_{12}^2}\right)} \tag{2-114}$$

Comparison of (2-114) with (2-86), which is the expression for the optimal covariance of the above example with just one measurement, suggests that this problem is equivalent to Example 2.1 with the two measurements replaced by a single equivalent measurement with variance

$$r = \frac{r_{11} + r_{22} - 2r_{12}}{r_{11}r_{12} - r_{12}^2} \tag{2-115}$$

Assuming R is not singular the denominator of (2-115) must be positive.

The derivative of r with respect to r_{12} is

$$\frac{dr}{dr_{12}} = \frac{2\left[\left(r_{11} - r_{12}\right)\left(r_{22} - r_{12}\right)\right]}{\left(r_{11}r_{22} - r_{12}^2\right)^2} \tag{2-116}$$

The only restriction on r_{12}, is that

$$r_{11}r_{22} > r_{12}^2 \tag{2-117}$$

Consider two cases. Suppose first that $r_{11} = r_{22} = 1$ and r_{12} is assumed to lie in the range

$$-\frac{1}{2} \leq r_{12} \leq \frac{1}{2} \tag{2-118}$$

Then (2-116) is always positive and the maximizing value of r_{12} is 1/2. If, however, $r_{11} = 4$ and $r_{22} = 1$ then

$$-\frac{3}{2} \leq r_{12} \leq \frac{3}{2} \tag{2-119}$$

satisfies (2-117). In this case dr/dr_{12} equals zero for $r_{12} = 1$. Since

$$\frac{dr}{dr_{12}} > 0 \quad \text{for} \quad r_{12} < 1 \tag{2-120}$$

and

$$\frac{dr}{dr_{12}} < 0 \quad \text{for} \quad r_{12} > 1 \tag{2-121}$$

it is clear that $r_{12} = 1$ is a maximum.

Example 2.4

This last example treats two first-order plants which are dynamically uncoupled. Coupling between the systems is introduced through the off-diagonal term in Q, which is uncertain. The system is shown in Fig. 2-4. The matrices for this example are:

$$F = \begin{bmatrix} -a & 0 \\ 0 & -b \end{bmatrix}; \quad G = H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{12} & q_{22} \end{bmatrix}; \quad R = \begin{bmatrix} r_{11} & 0 \\ 0 & r_{12} \end{bmatrix}$$

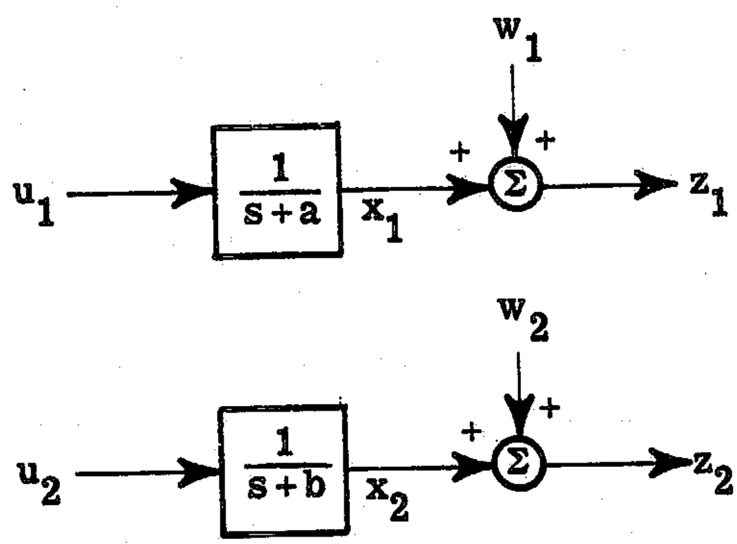


Figure 2-4 Block Diagram for Example 2.4

Kalman has discussed the complete solution of this example [16]. If det Q > 0 the steady-state optimal covariance of this system is given by these equations:

$$p_{11} = r_{11}\left(-a + \sqrt{\frac{q_{11}}{r_{11}} + a^2 - \frac{p_{12}^2}{r_{11}r_{12}}}\right) \tag{2-122}$$

$$p_{22} = r_{22}\left(-b + \sqrt{\frac{q_{22}}{r_{22}} + b^2 - \frac{p_{12}^2}{r_{11}r_{12}}}\right) \tag{2-123}$$

$$p_{12}^2 = \frac{r_{11}r_{22}}{\frac{q_{11}}{r_{11}} + \frac{q_{22}}{r_{22}} + a^2 + b^2 + 2\sqrt{\left(\frac{q_{11}}{r_{11}} + a^2\right)\left(\frac{q_{22}}{r_{22}} + b^2\right) - \frac{q_{12}^2}{r_{11}r_{22}}}} \tag{2-124}$$

From (2-124) it is clear that p_{12}^2 takes on its minimum value with respect to q_{12} when $q_{12} = 0$. Any non-zero value of q_{12} increases p_{12}^2. Eqs. (2-122) and (2-123) indicate that p_{11} and p_{22} are maximum when p_{12}^2 is a minimum. Thus the minimax for this example occurs when $q_{12} = 0$. The point to realize here is that statistical coupling of the two first-order plants through q_{12} generates non-zero correlation between z_1 and x_2 and between z_2 and x_1. Thus additional information about each of these states is available and the covariance of each drops accordingly.

2.8 Conclusion

In this chapter the value of the minimax filter return function, S_1, for uncertain noise statistics has been shown to be equal to the maximum value of the optimal filter return function over the uncertain parameter set. Furthermore, the minimax filter was shown to be the

Kalman filter for this maximizing set of parameters. Certain properties of the infinite time maximization problem were then developed. It was shown that the diagonal elements of the Q and R matrices are maximal and that the gradient of the optimal return function with respect to the uncertain parameters always exists. In Chapter IV computational aspects of the maximization will be discussed in greater detail.

CHAPTER III

THE MINIMAX SENSITIVITY FILTER FOR UNCERTAIN NOISE STATISTICS

3.1 Introduction

In this chapter some properties of minimax sensitivity filters for time invariant plants with constant but uncertain input and measurement noise covariance matrices are developed. Our attention will be limited to the infinite time problem. The infinite time filter takes the form

$$\dot{\overline{x}} = F\overline{x} + K(z - H\overline{x}) \tag{3-1}$$

where K is now a constant matrix. V_K is redefined as

$$V_K = \left\{K | R_e[\lambda_i(F - KH)] < 0, \quad 1 \le i \le n\right\} \tag{3-2}$$

That is, V_K is the set of all K such that (F - KH) is a stable matrix. From theorem 2.3

$$V_K \supset \overline{K}_o = \left\{K | K = \overline{K}_o \text{ for some } v \in V\right\} \tag{3-3}$$

Using eq. (2-16) the infinite time solution for M(t) is

$$\overline{M} = \int_0^\infty \Phi_K(\sigma)\left[KRK^T + GQG^T\right]\Phi_K^T(\sigma)\, d\sigma \tag{3-4}$$

$\overline{M}$ may also be found by equating $\dot{M}$ to zero in (2-15) and solving the resulting linear algebraic equation. $\overline{J}_M = \mathrm{tr}\,(W\overline{M})$ is not a function of P_o. Thus the infinite time absolute and relative performance sensitivities are functions of K and v only. They are

$$S^A(K, v) = \overline{J}_M(K, v) - \overline{J}_o(v) \tag{3-5}$$

$$S^R(K, v) = \frac{\overline{J}_M(K, v) - \overline{J}_o(v)}{\overline{J}_o(v)} \tag{3-6}$$

The infinite time minimax sensitivity filter is now defined.

Def. 3.1: The infinite time minimax sensitivity filter for uncertain but constant noise statistics is defined by that value of $K \in V_K$ denoted by $\hat{K}$ for which

$$\max_{v \in V} S(\hat{K}, v) = \min_{K \in V_K} \max_{v \in V} S(K, v) \tag{3-7}$$

where S may be S^R or S^A. Where confusion may arise a subscript will be placed on K to differentiate between the two.

Salmon has shown that algebraic performance sensitivities have no saddle point [10]. Thus results similar to those for the minimax filter are not available for minimax sensitivity filters. The regions in V_K and V where the minimax must lie, however, can be stated. In particular, it will be shown that the maximum of S^A or S^R is attained on the extreme points of V and that the minimizing value of K is an element of $\overline{K}_o$.

3.2 More on Convexity

A review of some additional definitions and properties of convex functions will aid in understanding the development of this chapter. Again much of this material can be found in the cited references on convexity [11], [12]. Where lacking, however, short proofs have been supplied.

Property 3.1: If X is a convex set and $x_i \in X$ for every i in some countable index set then the point x

$$x = \sum_i \alpha_i x_i; \quad 0 \le \alpha_i \le 1, \quad \sum_i \alpha_i = 1$$

is also in X.

This result is easily established by induction from Def. 2.1 and may be taken as an alternate definition of a convex set.

From definition 2.2 and property 3.1 we obtain immediately:

Property 3.2: If f(x) is a convex scalar function of x defined on a convex set X then

$$f\left(\sum_{i=1}^{r} \alpha_i x_i\right) \le \sum_{i=1}^{r} \alpha_i f(x_i)$$

where

$$x_i \in X, \quad 0 \le \alpha_i \le 1, \quad \sum_{i=1}^{r} \alpha_i = 1$$

and r is any positive integer.

Theorem 3.1: If f(x) is a continuous scalar function defined on a compact convex set X and if f(x) is strictly convex in X then f(x) assumes its minimum value at only one point in X.

Proof: Since X is compact, f(x) has at least one minimum. Suppose there are two minima at x_1 and x_2. Then by strict convexity

$$f\left(\frac{x_1 + x_2}{2}\right) < \frac{1}{2}f(x_1) + \frac{1}{2}f(x_2) = f(x_1) = f(x_2)$$

which says that f(x) evaluated at the midpoint between x_1 and x_2 is smaller than $f(x_1)$ or $f(x_2)$. Contradiction.

Def. 3.2: A point x in a convex set X is called an extreme point of X if there are no points x_1 and x_2 in X such that $x = \alpha x_1 + (1-\alpha)x_2$ for some α, $0 < \alpha < 1$, where $x_1 \neq x_2$.

A theorem will now be cited which is of central importance to the discussions of this chapter, namely the Krein-Millman theorem [12].

Theorem 3.2: A compact convex set is spanned by its extreme points. That is, every x in X can be represented in the form

$$x = \sum_{k=1}^{r} \alpha_k x_k; \quad 0 \leq \alpha_k \leq 1, \quad \sum_{k=1}^{r} \alpha_k = 1$$

where $x_1, \ldots, x_r$ are extreme points of X.

By way of comment, the extreme points of a convex set should not be confused with the basis vectors of a vector space. For example, a simple three-dimensional cube is a convex set and has eight (8) extreme points. These are the eight vertices of the cube. In general, the vertices of a convex polyhedron are also its extreme points. Every point inside the convex polyhedron can be written as a convex combination of its vertices. The set of all points on and within a sphere of radius a about the origin is a convex set. The extreme points of this set are the points x defined as

$$X_E = \{x \mid \|x\| = a\}$$

Observe that this set is not countable. We shall be concerned only with convex sets whose extreme points are finite or countably infinite.

<u>Theorem 3.3:</u> If f(x) is a convex scalar function of a vector x defined on a compact convex set X then

$$\max_{x \in X} f(x) = \max_{x \in X_E} f(x)$$

where X_E is the set of extreme points of X.

Proof: Suppose f(x) attains its maximum value at some interior point in X, say x_a. Then by theorem 3.2

$$x_a = \sum_i \alpha_i x_i; \quad 0 \le \alpha_i \le 1, \quad \sum_i \alpha_i = 1$$

$$x_i \in X_E$$

By convexity of f(x) in X one has

$$f(x_a) \le \sum_i \alpha_i f(x_i) \le \sum_i \alpha_i \max_i f(x_i)$$

or

$$f(x_a) \le \max_i f(x_i)$$

Since $f(x_a)$ is a maximum the above inequality can hold only with the equal sign. That is

$$f(x_a) = \max_{x \in X} f(x) = \max_{x_i \in X_E} f(x_i)$$

Theorem 3.3 simply says that any maximum of a convex function attained in the interior of X is also attained at one (or more) of its extreme points.

<u>Property 3.3:</u> Let $f_\nu(x)$ be a set of convex (strictly convex) functions on a compact convex set X. Then

$$\varphi(x) = \max_{\nu} f_{\nu}(x)$$

is convex (strictly convex) on X. Note: subscript ν denotes any index set.

3.3 Minimax Sensitivity Filters

Before presenting the central results of this chapter a few elementary properties of infinite time filter performance measures and performance sensitivity measures of interest are developed in the following lemmas:

Lemma 3.1: The optimal return function $\overline{J}_0(v)$ is concave in V.

Proof: For every $v \in V$ there exists a $\overline{K}_0(v)$ such that

$$\overline{J}_M\left[\overline{K}_0(v), v\right] = \overline{J}_0(v) \tag{3-8}$$

Now V is convex and $J_M(K, v)$ is linear in V so that (3-8) can be written

$$\overline{J}_0(v) = \overline{J}_M(\overline{K}_0, v) = \overline{J}_M\left[\overline{K}_0, \alpha v_1 + (1-\alpha)v_2\right]; \quad 0 \le \alpha \le 1 \tag{3-9}$$

or

$$\overline{J}_0(v) = \alpha\overline{J}_M(\overline{K}_0, v_1) + (1-\alpha)\overline{J}_M(\overline{K}_0, v_2) \tag{3-10}$$

but $\overline{J}_M(K, v) \ge \overline{J}_0(v) \ge 0$ for every v. Thus from (3-10)

$$\overline{J}_0(v) \ge \alpha\overline{J}_0(v_1) + (1-\alpha)\overline{J}_0(v_2) \tag{3-11}$$

Lemma 3.2: The absolute performance sensitivity $S^A(K, v)$ is convex in V.

Proof: for $v = \alpha v_1 + (1-\alpha)v_2$ one has

$$\bar{J}_M(K, v) = \alpha \bar{J}_M(K, v_1) + (1-\alpha)\bar{J}_M(K, v_2) \tag{3-12}$$

and from lemma 3.1

$$\bar{J}_o(v) \geq \alpha \bar{J}_o(v_1) + (1-\alpha)\bar{J}_o(v_2) \tag{3-13}$$

Subtracting (3-13) from (3-12) yields

$$\bar{J}_M(K, v) - \bar{J}_o(v) \leq \alpha \left[\bar{J}_M(K, v_1) - \bar{J}_o(v_1)\right] + (1-\alpha)\left[\bar{J}_M(K, v_2) - \bar{J}_o(v_2)\right]$$

or

$$S^A(K, v) \leq \alpha S^A(K, v_1) + (1-\alpha) S^A(K, v_2) \tag{3-14}$$

which was to be shown.

Lemma 3.3: The relative performance sensitivity $S^R(K, v)$ is convex in V.

Proof: Dividing (3-14) by (3-11) yields

$$\frac{S^A(K, v)}{\bar{J}_o(v)} \leq \frac{\alpha S^A(K, v_1) + (1-\alpha) S^A(K, v_2)}{\alpha \bar{J}_o(v_1) + (1-\alpha)\bar{J}_o(v_2)}$$

$$\leq \alpha \frac{S^A(K, v_1)}{\bar{J}_o(v_1)} + (1-\alpha) \frac{S^A(K, v_2)}{\bar{J}_o(v_2)} \tag{3-15}$$

or

$$S^R(K, v) \leq \alpha S^R(K, v_1) + (1-\alpha)S^R(K, v_2) \tag{3-16}$$

which is the desired result.

Lemma 3.4: The performance sensitivities are strictly convex in V_K for every $v \in V$.

Proof: In Chapter II $J_M(K, v)$ was shown to be strictly convex in V_K for every $v \in V$. Then for $K = \alpha K_1 + (1-\alpha)K_2$

$$\bar{J}_M(K, v) < \alpha \bar{J}_M(K_1, v) + (1-\alpha)\bar{J}_M(K_2, v) \tag{3-17}$$

Subtracting $\bar{J}_o(v)$ from both sides of (3-17) yields

$$\bar{J}_M(K, v) - \bar{J}_o(v) < \alpha\left[\bar{J}_M(K_1, v) - \bar{J}_o(v)\right] + (1-\alpha)\left[\bar{J}_M(K_2, v) - \bar{J}_o(v)\right]$$

or

$$S^A(K, v) < \alpha S^A(K_1, v) + (1-\alpha) S^A(K_2, v) \tag{3-18}$$

which is the first desired result. Dividing (3-18) by $J_o(v)$ then gives

$$S^R(K_1, v) < \alpha S^R(K_1, v) + (1-\alpha) S^R(K_2, v) \tag{3-19}$$

We come now to the central results of this chapter. The first is rather straightforward as the following lemma reveals:

Lemma 3.5: Let S(K, v) denote a performance sensitivity and let V_E denote the extreme points of V. Denote by $\tilde{K}$ the value of K for which

$$\max_{v \in V_E} S(\tilde{K}, v) = \min_{K \in V_K} \max_{v \in V_E} S(K, v)$$

then $\tilde{K} = \hat{K}$.

Proof: Since S^A and S^R are convex in V we have by theorem 3.3

$$\max_{v \in V} S(K, v) = \max_{v \in V_E} S(K, v) \tag{3-20}$$

Taking the min of both sides of (3-20) with respect to $K \in V_K$ yields

$$\min_{K \in V_K} \max_{v \in V} S(K, v) = \min_{K \in V_K} \max_{v \in V_E} S(K, v) \tag{3-21}$$

or

$$\max_{v \in V} S(\hat{K}, v) = \max_{v \in V_E} S(\tilde{K}, v) = \max_{v \in V} S(\tilde{K}. v) \tag{3-22}$$

which together with lemma 3.4 and theorem 3.1 implies $\tilde{K} = \hat{K}$.

Lemma 3.6: The value of K which defines the minimax filter is unique.

Proof: This follows immediately from theorem 3.1 since by property 3.3 $\max_{v \in V} S(K, v)$ is strictly convex in V_K.

Lemma 3.6 tells us that the minimax sensitivity filter for uncertain Q and R is unique. Lemma 3.5 tells us that in searching for the smallest maximum we may restrict our attention to the extreme points of V. This is a finite set of points. The last theorem to be presented in this section tells us where to look for $\hat{K}$.

Theorem 3.4: The minimax sensitivity filter for constant but uncertain Q and R is optimal for some $v \in V$, that is

$$\hat{K} \in \overline{K}_o$$

Discussion: The theorem has great geometric appeal. Referring to Fig. 2.1, because $\overline{J}_M(K, v)$ is linear in V the surface representing $\overline{J}_M$ is a plane in three-dimensional space. Since the surface representing $\overline{J}_o(v)$ is concave in $q \times r$ it seems quite reasonable that the maximum deviation between $\overline{J}_M$ and $\overline{J}_o$ can only be minimized when the $\overline{J}_M$ plane is tangent to the $\overline{J}_o$ surface.

The proof of this theorem will proceed as follows: It will be shown that if $\hat{K} \notin \overline{K}_o$ a direction of motion in V_K, say ΔK, always exists such that

$$\overline{J}_M(\hat{K} + \epsilon\Delta K, v) < \overline{J}_M(\hat{K}, v); \quad \forall v \in V \tag{3-23}$$

where ϵ is an appropriate small positive constant. But (3-23) implies that

$$S(\hat{K}, v) > S(\hat{K} + \epsilon \Delta K, v); \quad \forall v \in V$$

and thus

$$\max_{v \in V} S(\hat{K}, v) > \max_{v \in V} S(\hat{K} + \epsilon \Delta K, v) \tag{3-24}$$

and therefore $\hat{K}$ cannot be the minimax sensitivity filter gain. Indeed the condition that no direction in V_K can be found which simultaneously reduces $J_M(K, v)$ for every $v \in V$ is precisely the condition that $K \in \overline{K}_o$.

By theorem 3.2 any point in V can be written as a convex combination of points in V_E. We have, therefore, that

$$\overline{J}_M(K, v) = \overline{J}_M(K, \sum_{i=1}^{r} \alpha_i v_i); \quad v_i \in V_E \tag{3-25}$$

Since J_M is linear in V, (3-25) may be written as

$$\overline{J}_M(K, v) = \sum_{i=1}^{r} \alpha_i \overline{J}_M^i(K, v_i) \tag{3-26}$$

The $\overline{J}_M^i(K, v_i)$, $i = 1, \ldots, r$, constitute a set of basis functionals for the functional $\overline{J}_M(K, v)$. Since the α_i are all positive we need only show that for $K \notin \overline{K}_o$ a ΔK always exists such that

$$\overline{J}_M^i(K + \epsilon \Delta K, v_i) < \overline{J}_M^i(K, v_i); \quad \forall i \tag{3-27}$$

A heuristic proof will first be given limiting V_K to a three-dimensional space where pictures can be drawn. From this proof the analytic condition for the general proof will be geometrically apparent. Note first, however, from the derivation of the Kalman filter [eq. (5-15)] that

$$\nabla_K \bar{J}_M(K, v)\Big|_{\substack{K=K_1 \\ v=v_1}} = 0 \Rightarrow \bar{J}_M(K_1, v_1) = \bar{J}_o(v_1) \qquad (3\text{-}28)$$

or equivalently $K_1 \in \bar{\mathbf{K}}_o$.

Proof I: The proof will proceed by induction on the basis functionals $\bar{J}_M^i(K, v_i)$

1. Assume that $\hat{K} \notin \bar{\mathbf{K}}_o$ then $\nabla_K \bar{J}_M(\hat{K}, v_1) \neq 0$ and $\Delta K = -\nabla_K \bar{J}_M(\hat{K}, v_1)$ is an appropriate direction to reduce $\bar{J}_M^1(\hat{K}, v_1)$.[1]
2. Consider $\bar{J}_M^1$ and $\bar{J}_M^2$. $\nabla_K \bar{J}_M^1(\hat{K}, v_1)$ and $\nabla_K \bar{J}_M^2(\hat{K}, v_2)$ define two planes in V_K passing through $\hat{K}$. They also define two open half spaces in which the directional derivative of $\bar{J}_M^1$ or $\bar{J}_M^2$ is negative. If the intersection of these two half spaces is non-void a direction in V_K can be found in which $\bar{J}_M^1$ and $\bar{J}_M^2$ can both be reduced. The intersection is void if and only if

---

[1] For convenience the notation

$$\nabla_K \bar{J}_M(\hat{K}, v) = \nabla_K \bar{J}_M(K, v)\Big|_{K=\hat{K}}$$

is used throughout this section.

$$\nabla_K \bar{J}_M^1(\hat{K}, v_1) = -\beta \nabla_K \bar{J}_M^2(\hat{K}, v_2) \tag{3-29}$$

where β is some positive constant. Now consider a point $\tilde{v} = \alpha v_1 + (1-\alpha)v_2$ and the functional

$$\bar{J}_M(\hat{K}, \tilde{v}) = \alpha \bar{J}_M^1(\hat{K}, v_1) + (1-\alpha)\bar{J}_M^2(\hat{K}, v_2) \tag{3-30}$$

Then

$$\nabla_K \bar{J}_M(\hat{K}, \tilde{v}) = \alpha \nabla_K \bar{J}_M^1(\hat{K}, v_1) + (1-\alpha)\bar{J}_M^2(\hat{K}, v_2) \tag{3-31}$$

but by (3-29)

$$\nabla_K \bar{J}_M(\hat{K}, \tilde{v}) = [\alpha - \beta(1-\alpha)] \nabla_K \bar{J}_M^1(\hat{K}, v_1) \tag{3-32}$$

Setting $\alpha - \beta(1-\alpha) = 0$ yields

$$\alpha_\beta = \frac{\beta}{\beta+1} < 1 \tag{3-33}$$

which is a legitimate value of α. Thus for $\tilde{v} = \alpha_\beta v_1 + (1-\alpha_\beta)v_2$ one has

$$\nabla_K \bar{J}_M(\hat{K}, \tilde{v}) = 0 \tag{3-34}$$

but (3-34) implies that $\hat{K} \in \bar{K}_o$. Contradiction. That is, (3-29) is not valid and the intersection is not void. At this point we know that no two gradient hyperplanes of two basis functionals are parallel when $\hat{K} \notin \bar{K}_o$.

3. A third basis functional is now added. It is clear from the foregoing that all three gradient planes intersect at $\hat{K}$. If they intersect only at that point the intersection of the three negative half spaces forms a convex polyhedral cone with apex at $\hat{K}$ and any motion into that cone from $\hat{K}$ will simultaneously reduce all three basic functionals.

We need only consider the case where all three gradient planes meet in a line of intersection. Even in this instance it is possible for the intersection of the three negative half spaces to be non-void. However, if $\nabla_K \bar{J}^3_M(K, v_3)$ at $\hat{K}$ lies in the convex span of the negatives of $\nabla_K \bar{J}^1_M(\hat{K}, v_1)$ and $\nabla_K \bar{J}^2_M(\hat{K}, v_2)$ the intersection of all three negative half spaces is void. This condition is diagrammed in Fig. 3-1. Analytically this is equivalent to

$$\nabla_K \bar{J}^3_M(\hat{K}, v_3) = -t\left[\alpha \nabla_K \bar{J}^1_M(\hat{K}, v_1) + (1-\alpha)\nabla_K \bar{J}^2_M(\hat{K}, v_2)\right] \quad (3\text{-}35)$$

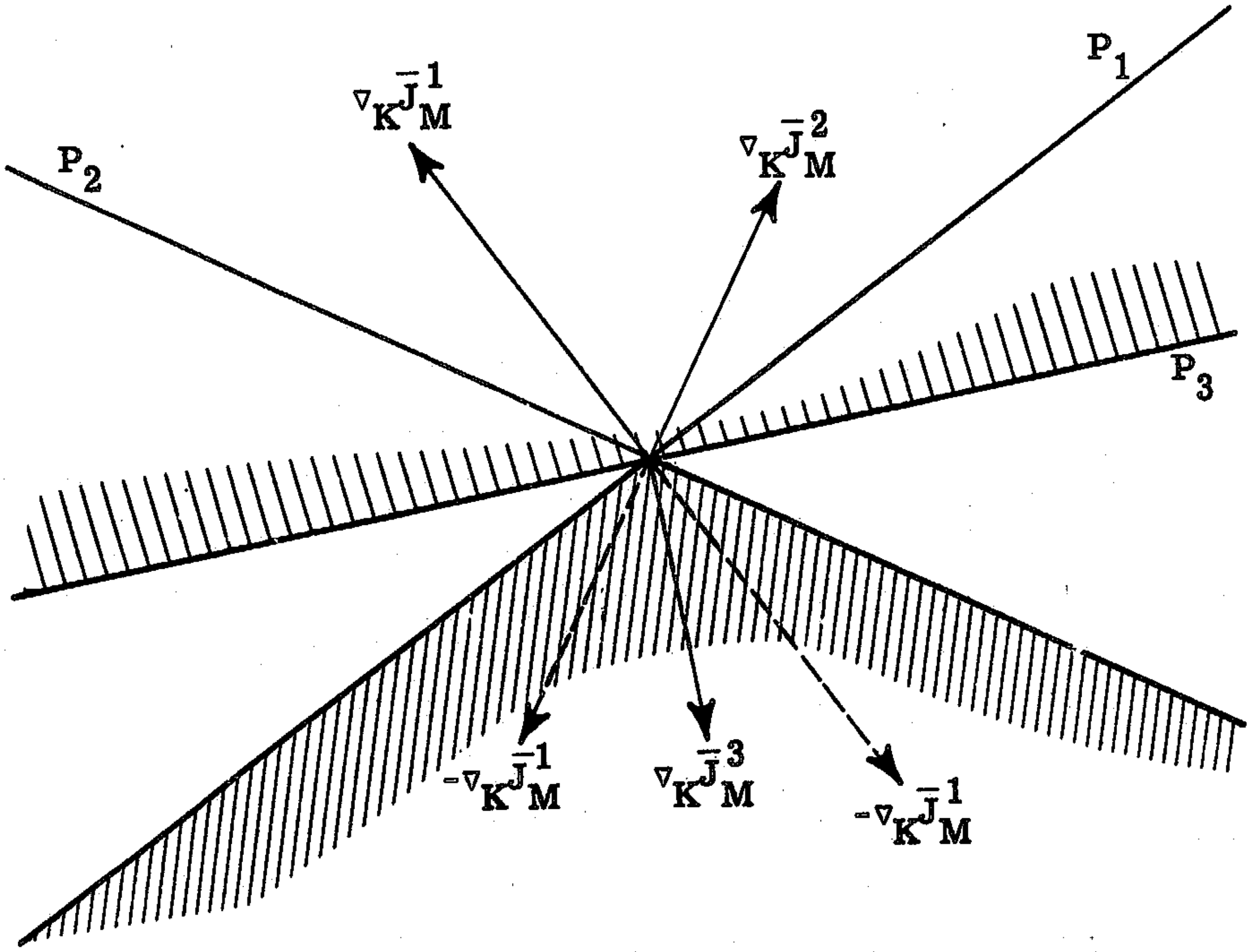
where t is some positive constant.

Now consider the point

$$\tilde{v} = \sum_{i=1}^{3} \gamma_i v_i \quad (3\text{-}36)$$

and the performance index

$$\bar{J}_M(K, \tilde{v}) = \sum_{i=1}^{3} \gamma_i \bar{J}^i_M(K, v_i) \quad (3\text{-}37)$$



P_1 plane defined by $\nabla_K \bar{J}_M^1$

P_2 plane defined by $\nabla_K \bar{J}_M^2$

P_3 plane defined by $\nabla_K \bar{J}_M^3$

common region of negative directional derivative for $\bar{J}_M^1$ and $\bar{J}_M^2$

region of negative directional derivative for $\bar{J}_M^3$

Figure 3-1 Gradient Planes in Heuristic Proof of Theorem 3.4

Then

$$\nabla_K \bar{J}_M(K, \bar{v}) = \sum_{i=1}^{3} \gamma_i \nabla_K \bar{J}^i_M(K, v_i) \tag{3-38}$$

Evaluating (3-38) at $\hat{K}$ and using (3-35) one obtains

$$\begin{aligned} \nabla_K \bar{J}_M(\hat{K}, v) &= \gamma_1 \nabla_K \bar{J}^1_M(\hat{K}, v_1) + \gamma_2 \nabla_K \bar{J}^2_M(\hat{K}, v_2) \\ &\quad - \gamma_3 t\left[\alpha \nabla_K \bar{J}^1_M(\hat{K}, v_1) + (1-\alpha) \nabla_K \bar{J}^2_M(\hat{K}, v_2)\right] \end{aligned} \tag{3-39}$$

The expression equals zero if

$$\gamma_1 - \gamma_3 t\alpha = 0 \tag{3-40}$$

and

$$\gamma_2 - \gamma_3 t(1-\alpha) = 0 \tag{3-41}$$

Eqs. (3-40) and (3-41) together with $\gamma_1 + \gamma_2 + \gamma_3 = 1$ give us a system of three linear equations for the γ_i. This system of equations always has a solution.[2] Furthermore, from eqs. (3-40) and (3-41) it is clear that γ_1 and γ_2 both have the same sign as γ_3 and thus all must be positive and satisfy the inequality constraint $0 \leq \gamma_i \leq 1$. Thus a $v \in V$ has been found for which $\nabla_K J_M(\hat{K}, v) = 0$ is zero implying that $\hat{K} \in \bar{K}_o$. Contradiction. That is, the assumption that

[2]See proof of lemma 3.7.

no direction of motion in V_K exists which will simultaneously reduce the first three basis functionals when $\hat{K} \notin \bar{K}_o$ leads to the conclusion that $\hat{K} \in \bar{K}_o$.

This heuristic proof can be continued until all basis functionals are exhausted, but the point has been made. The formal proof of theorem 3.4 is a direct consequence of the next lemma.

Lemma 3.7: Let

$$\bar{J}^i_M(K, v_i); \quad v_i \in V_E, \quad i = 1, \ldots r$$

represent the set of basis functionals for the performance index $J_M(K, v)$, and let K_1 be an element of $V_K \setminus \bar{K}_o$. Then there exists a direction ΔK in V_K such that

$$\bar{J}^i_M(K_1 + \epsilon \Delta K, v_i) < \bar{J}^i_M(K_1, v_i); \quad \forall i \tag{3-43}$$

and ϵ an appropriate positive constant.

Proof:

1. Recall that when $K \in V_K \setminus \bar{K}_o$, $\nabla_K \bar{J}_M(K, v) \neq 0$ for any $v \in V$. Now by selecting $\Delta K = -\nabla_K \bar{J}^1_M(K_1, v_1)$ the first basis functional can be reduced in value.
2. Assume (3-43) is true for $i = 1, \ldots, n$ and denote by a_i the gradient $\nabla_K \bar{J}^i_M(K_1, v_i)$. Then there exists a convex polyhedral cone C_n, defined as follows:

$$C_n = \left\{\Delta K \mid \langle a_i, \Delta K\rangle < 0, \quad i = 1,\ldots,n\right\} \tag{3-44}$$

To establish the next step in the induction we must show that the intersection of the negative half space

$$S = \left\{\Delta K_{n+1} \mid \langle a_{n+1}, \Delta K_{n+1}\rangle < 0\right\} \tag{3-45}$$

with C_n is non-void. Assuming the contrary, observe that the condition $C_n \cap S = \emptyset$ is precisely the condition that

$$a_{n+1} = -t\left(\sum_{i=1}^{n} \alpha_i a_i\right); \quad 0 \le \alpha_i \le 1, \quad \sum_{i=1}^{n} \alpha_i = 1 \tag{3-46}$$

i.e., a_{n+1} lies in the negative convex span of $a_1, \ldots, a_n$. When (3-46) is true one has from (3-44) for any $\Delta K \in C_n$

$$\langle a_{n+1}, \Delta K\rangle = -t\sum_{i=1}^{n} \alpha_i \langle a_i, \Delta K\rangle > 0 \tag{3-47}$$

Thus no small motion in C_n at K_1 will reduce $\bar{J}_M^{n+1}(K, v_{n+1})$. Now consider the point $\tilde{v} \in V$

$$\tilde{v} = \sum_{i=1}^{n+1} \gamma_i v_i; \quad v_i \in V_E, \quad 0 \le \gamma_i \le 1, \quad \sum_{i=1}^{n+1} \gamma_i = 1 \tag{3-48}$$

and the performance functional

$$\bar{J}_M(K, \tilde{v}) = \sum_{i=1}^{n+1} \gamma_i \bar{J}_M^i(K, v_i) \tag{3-49}$$

Then

$$\nabla_K \bar{J}_M(K, \tilde{v}) = \sum_{i=1}^{n+1} \gamma_i \nabla_K \bar{J}_M^i(K, v_i) \tag{3-50}$$

and using (3-46) one obtains

$$\nabla_K \bar{J}_M(K, \tilde{v}) = \sum_{i=1}^{n} \gamma_i \nabla_K \bar{J}_M^i(K, v_i) - t\gamma_{n+1} \sum_{i=1}^{n} \alpha_i \nabla_K \bar{J}_M^i(K, v_i) \tag{3-51}$$

or

$$\nabla_K \bar{J}_M(K, \tilde{v}) = \sum_{i=1}^{n} \left(\gamma_i - \alpha_i t \gamma_{n+1}\right) \nabla_K \bar{J}_M^i(K, v_i) \tag{3-52}$$

This gradient can be made zero at K_1 by equating all coefficients to zero and invoking the constraint on the γ_i. This leads to a set of $n+1$ linear equations, in $n+1$ unknowns, the γ_i.

$$\begin{aligned} \gamma_1 - \alpha_1 t \gamma_{n+1} &= 0 \\ \gamma_2 - \alpha_2 t \gamma_{n+1} &= 0 \\ \vdots \qquad & \quad \vdots \\ \gamma_n - \alpha_n t \gamma_{n+1} &= 0 \\ \gamma_1 + \gamma_2 + \dots + \gamma_{n+1} &= 1 \end{aligned} \tag{3-53}$$

From the first n of these equations it is clear that all γ_i have the same sign. The (n+1)th equation requires that they all be positive. Putting the last equation into the first n equations yields

$$\begin{aligned}
(1+\alpha_1 t)\gamma_1 + \alpha_1 t\gamma_2 + \dots + \alpha_1 t\gamma_n &= \alpha_1 t \\
\alpha_2 t\gamma_1 + (1+\alpha_2 t)\gamma_2 + \dots + \alpha_2 t\gamma_n &= \alpha_2 t \\
\vdots \qquad\qquad & \qquad \vdots \\
\alpha_n t\gamma_1 + \quad \dots \quad + (1+\alpha_n t)\gamma_n &= \alpha_n t
\end{aligned} \tag{3-54}$$

Summing these equations one obtains

$$(1+t)\sum_{i=1}^{n} \gamma_i = t$$

or

$$\sum_{i=1}^{n} \gamma_i = \frac{t}{1+t} \tag{3-55}$$

so that

$$\gamma_{n+1} = 1 - \sum_{i=1}^{n} \gamma_i = \frac{1}{1+t} < 1 \tag{3-56}$$

and

$$\gamma_i = \frac{\alpha_i t}{1+t} < 1, \quad i = 1,\dots,n \tag{3-57}$$

Thus an appropriate set of γ_i exists, i.e., a set satisfying (3-48), and therefore there exists a $\tilde{v} \epsilon V$ such that (3-52) is zero. That is

$$\nabla_K \bar{J}_M(K, \tilde{v})\Big|_{K=K_1} = 0 \tag{3-58}$$

which implies $K_1 \epsilon \bar{K}_0$. Contradiction.

As the preliminary discussion indicates, proof of theorem 3.5 follows immediately from lemma 3.7.

Salmon, in treating the minimax control problem has shown by example that the minimax sensitivity control is not necessarily an optimal control [10]. It has been shown here, however, that when the optimal return function is concave in V and the suboptimal return function is linear in V, the minimax control (or filter gain) is indeed optimal for some $v \epsilon V$.

Recapping the case for constant but uncertain Q and R matrices, the minimax sensitivity filter was shown to be unique (lemma 3.6). Furthermore, in searching for the minimax sensitivity filter we may by lemma 3.5 and theorem 3.4 restrict our attention to the extreme points of V and the set of optimal gains $\bar{K}_0$.

A simple example is now presented.

Example 3.1

The minimax absolute sensitivity filter will be determined for example 2.1 and its error performance compared with that of the minimax filter. Numerical values will be selected for the ranges of q and r to make the comparison explicit. Let

$$0 \le q \le 1 \tag{3-59a}$$

and

$$0 \le r \le 1 \tag{3-59b}$$

Since S^R is infinite when $r = 0$ or $q = 0$ the minimax relative sensitivity filter does not exist for the selected ranges of q and r. The performance index for filter (2-79) is

$$\bar{J}_M(k, r, q) = \frac{k^2 r + q}{2(1 + k)} \tag{3-60}$$

At the $q \times r$ extreme points (0, 1) and (1, 0) $\bar{J}_M$ is equal to S^A since J_o is zero. S^A will have the same numerical value at these two points if

$$k = 1 \tag{3-61}$$

and for this value of k

$$S^A = 0.25 \tag{3-62}$$

For $k = 1$ the values of S^A at (0, 0) and (1, 1) are respectively 0 and 0.086. Since a value of k has been found which maximizes S^A at two distinct points in $q \times r$ that k must yield the minimax sensitivity filter [10].

A numerical comparison of the S_1 and S_2 filters is contained in Table 3.1.

Table 3.1

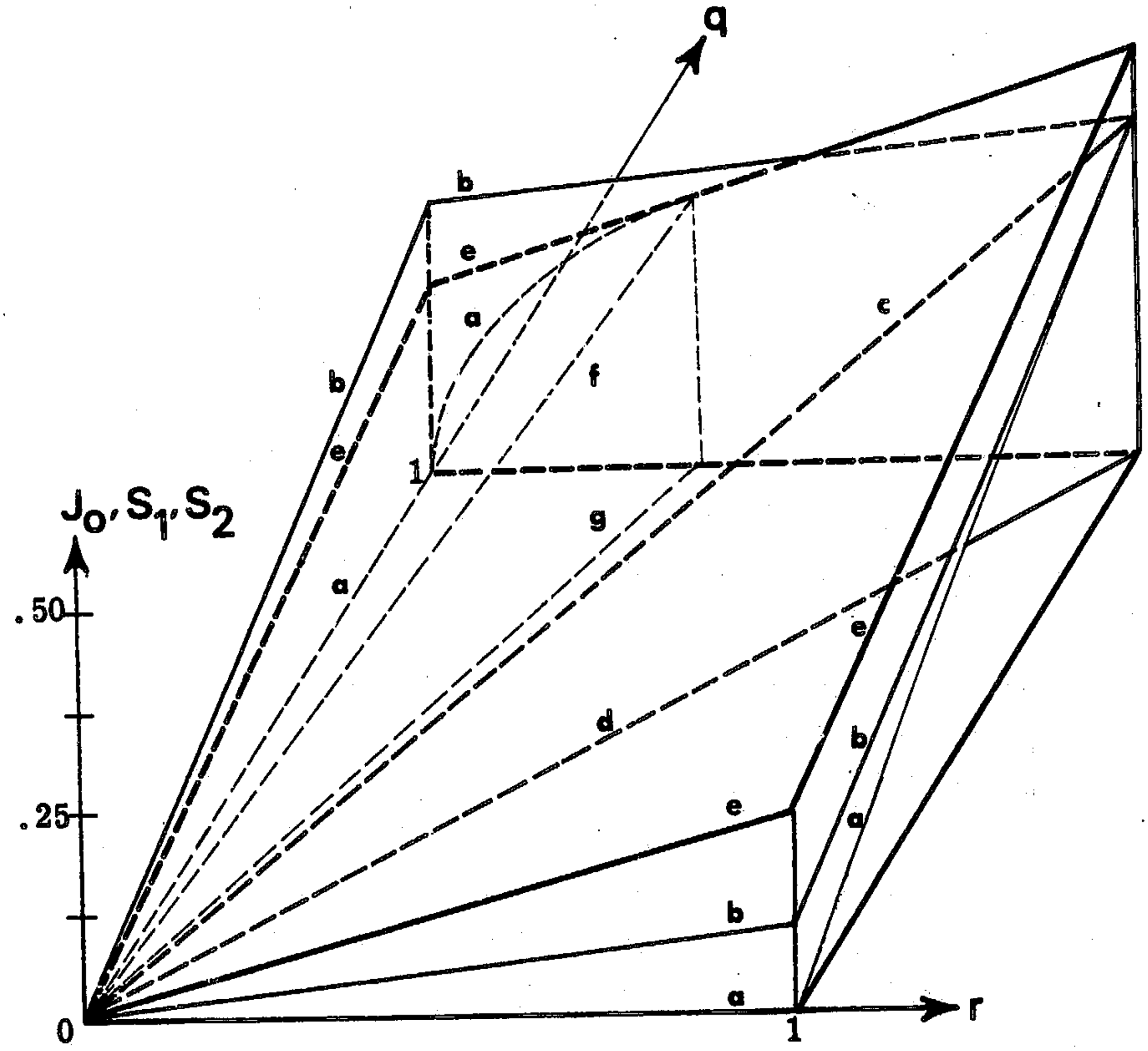
S_1, S_2 Filter Comparison

EXTREME POINT	MEAN SQUARE ERROR				
q, r	OPT	S_1	S_2	Δ_1	Δ_2
0, 0	0	0	0	0	0
0, 1	0	.068	.25	.068	.250
1, 0	0	.353	.25	.353	.250
1, 1	.414	.414	.500	0	.086
Max Error	.414	.414	.500		
Max Δ	0	.353	.250		

It is seen that the S_2 filter provides a smaller maximum deviation from optimality at the expense of greater absolute error whereas the opposite is true of the S_1 filter. The optimal, S_1 and S_2 filter error surfaces are shown in Fig. 3-2. The value of k in (3-60) corresponds to a q/r ratio of 3. Thus the S_2 filter error plane intersects the optimal error surface along the line q = 3r.

3.4 The Extreme Point Problem

The theory of the preceding section glosses over a very practical problem. How does one find the extreme points of V? Also unanswered is the question raised in Chapter II; namely, given fixed diagonal terms for Q and R, does a simple technique exist for determing



a	J_o (optimal) error surface	d	projection of c on q × r
b	S_1 filter error plane	e	S_2 filter error plane
c	S_1, J_o line of contact	f	S_2, J_o line of contact

g projection of f on q × r

Figure 3-1 S_1, S_2, and Optimal Error Surfaces for Example 3.1

bounds on the off-diagonal terms such that within those bounds R and Q are respectively always positive definite and positive semi-definite?

With regard to the first question, consider a symmetric $n \times n$ positive semi-definite matrix. The principle minors determinant test for positive semi-definiteness provides n generally nonlinear constraints between the $N = n(n + 1)/2$ distinct elements of the matrix [15]. Any $N - n$ terms can be set to one of their extreme values and the extremes of the remaining n terms computed subject to these constraints. A given set of $N - n$ elements has 2^{N-n} extreme points, and for each of these points 2^n extremes can be found in the remaining n elements. This process can be repeated as many times as $N - n$ elements can be selected from N elements. Thus a total of

$$\mathbf{N} = \frac{N!}{n!(N-n)!} 2^N$$

tests for extreme points must be made. For a 4×4 matrix $\mathbf{N}$ exceeds 2×10^4! Because of the additional constraints on the diagonal elements, not all of the extreme points determined in this manner will be distinct, but even a simple 2×2 positive definite matrix ($\mathbf{N} = 24$) can have up to twenty distinct extreme points.

In this section a method will be presented for generating the extreme points of subsets of V_Q and V_R by inspection. This simplified technique will naturally limit the sets of matrices one can consider, but it is clearly required if the problem of generating extreme points is to be reduced to manageable proportions.

First observe that in purely diagonal matrices there are no constraints between terms. Such matrices simply have 2^n vertices. Now for any positive semi-definite 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}.$$

one must have

$$a_{11} \geq 0 \tag{3-63}$$

$$a_{11}a_{22} - a_{12}^2 \geq 0 \tag{3-64}$$

which together imply that $a_{22} \geq 0$. If A is to be positive definite these inequalities must hold strictly. Note that from a statistical standpoint inequality (3-64) limits the maximum cross-correlation between two random processes. Constraints (3-63) and (3-64) are easily verified by inspection or simple calculation. The technique to be presented will use only the above two inequalities.

Look at the second order positive semi-definite quadratic form

$$Q(x_1, x_2) = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2 \geq 0 \tag{3-65}$$

Then since $a_{11} \geq 0$ and $a_{22} \geq 0$, (3-65) is equivalent to

$$a_{11}x_1^2 + a_{22}x_2^2 \geq 2a_{12}x_1x_2; \quad \forall x_1, x_2 \tag{3-66}$$

Now any value of a_{12} satisfying (3-64) must satisfy (3-66). This is seen from the following inequality

$$\left(a_{11}x_1^2 + a_{22}x_2^2\right)^2 - 4a_{11}a_{22}x_1^2x_2^2 = \left(a_{11}x_1^2 - a_{22}x_2^2\right)^2 \geq 0 \tag{3-67}$$

which implies that

$$\left(a_{11}x_1^2 + a_{22}x_2^2\right)^2 \geq 4a_{11}a_{22}x_1^2x_2^2 \tag{3-68}$$

Inequality (3-64) in eq. (3-68) now yields

$$\left(a_{11}x_1^2 + a_{22}x_2^2\right)^2 \geq 4a_{12}^2x_1^2x_2^2$$

or

$$a_{11}x_1^2 + a_{22}x_2^2 \geq 2a_{12}x_1x_2 \tag{3-69}$$

which is inequality (3-66).

The quadratic form associated with an $n \times n$ symmetric matrix, A, is

$$Q(x) = \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} a_{ij}x_ix_j \tag{3-70}$$

Denote by $\bar{a}_{ij}$ any value of a_{ij} satisfying the cross-correlation constraint

$$a_{ii}a_{jj} \geq a_{ij}^2 \tag{3-71}$$

Then

$$a_{ii}x_i^2 + a_{jj}x_j^2 \geq 2\bar{a}_{ij}x_i x_j \tag{3-72}$$

and

$$\sum_{i=1}^{n} \sum_{j=i+1}^{n} \left(a_{ii}x_i^2 + a_{jj}x_j^2\right) \geq 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} \bar{a}_{ij}x_i x_j \tag{3-73}$$

Performing the indicated summation on the right yields

$$(n-1)\sum_{i=1}^{n} a_{ii}x_i^2 \geq 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} \bar{a}_{ij}x_i x_j$$

or

$$\sum_{i=1}^{n} a_{ii}x_i^2 \geq 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{\bar{a}_{ij}}{n-1} x_i x_j \tag{3-74}$$

Thus (3-70) will always be positive definite (or semi-definite) if

$$a_{ij} \leq \frac{\bar{a}_{ij}}{n-1} \tag{3-75}$$

The use of eqs. (3-75) and (3-63) is best illustrated by some examples.

Example 3.2

Consider first a 2×2 positive definite matrix in which

$$4 \le a_{11} \le 9 \quad \text{and} \quad 1 \le a_{22} \le 4 \tag{3-76}$$

Let the second order constraint required to insure positive definiteness be

$$\bar{\rho}_{12}^2 = \max\left(\frac{a_{12}^2}{a_{11}a_{22}}\right) = \frac{1}{4} \tag{3-77}$$

where $\bar{\rho}_{12}$ is the maximum second-order cross-correlation to be considered. The extreme points of the set defined by (3-76) and (3-77) are then simply:

$$\begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 9 & -3 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} 9 & \frac{3}{2} \\ \frac{3}{2} & 1 \end{bmatrix} \begin{bmatrix} 9 & -\frac{3}{2} \\ -\frac{3}{2} & 1 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & -1 \\ -1 & 1 \end{bmatrix}$$

Example 3.3

Denote by **A** the set of all positive semi-definite 3×3 matrices satisfying the first and second order constraints

$$0 \le a_{11} \le 9; \quad 0 \le a_{22} \le 4; \quad 0 \le a_{33} \le 1 \tag{3-78}$$

and

$$\bar{\rho}_{12} = \bar{\rho}_{13} = \bar{\rho}_{23} = 1 \tag{3-79}$$

Since it must be possible to generate any value of one diagonal term with all other elements equal to zero, the matrices

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 9 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

are clearly extreme points of this set. The next set of extreme points generates all possible (2×2) positive semi-definite submatrices of $\mathbf{A}$ in combination with an isolated third diagonal term. This set has twelve points, viz.:

$$\begin{bmatrix} 9 & -6 & 0 \\ -6 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 9 & -6 & 0 \\ -6 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 9 & 6 & 0 \\ 6 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 9 & 6 & 0 \\ 6 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 9 & 0 & -3 \\ 0 & 0 & 0 \\ -3 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 9 & 0 & -3 \\ 0 & 4 & 0 \\ -3 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 9 & 0 & 3 \\ 0 & 0 & 0 \\ 3 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 9 & 0 & 3 \\ 0 & 4 & 0 \\ 3 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & -2 \\ 0 & -2 & 1 \end{bmatrix} \quad \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & -2 \\ 0 & -2 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{bmatrix} \quad \begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 1 \end{bmatrix}$$

Observe, for instance that

$$\begin{bmatrix} 9 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

is not an extreme point of $\mathbf{A}$ since it can be written as a convex combination of the first and third matrices above. The last set of extreme points of $\mathbf{A}$ are generated by employing inequality (3-75) for n = 3 and

$$\bar{a}_{12} = 6, \quad \bar{a}_{13} = 3 \quad \text{and} \quad \bar{a}_{23} = 2 \tag{3-80}$$

These points are

$$\begin{bmatrix} 9 & -3 & -1.5 \\ -3 & 4 & -1 \\ -1.5 & -1 & 1 \end{bmatrix} \begin{bmatrix} 9 & 3 & 1.5 \\ 3 & 4 & 1 \\ 1.5 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & -3 & 1.5 \\ -3 & 4 & 1 \\ 1.5 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & 3 & -1.5 \\ 3 & 4 & -1 \\ -1.5 & -1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 9 & -3 & 1.5 \\ -3 & 4 & 1 \\ 1.5 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & 3 & -1.5 \\ 3 & 4 & 1 \\ -1.5 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & -3 & -1.5 \\ -3 & 4 & 1 \\ -1.5 & 1 & 1 \end{bmatrix} \begin{bmatrix} 9 & -3 & 1.5 \\ -3 & 4 & -1 \\ 1.5 & -1 & 1 \end{bmatrix}$$

for a total of twenty-four extreme points.

Example 3.4:

Let $\mathbf{B}$ be the set of all positive definite 3×3 matrices with constraints on the diagonal elements given by

$$4 \le b_{11} \le 16; \quad 1 \le b_{22} \le 9 \quad \text{and} \quad 1 \le b_{33} \le 4 \tag{3-81}$$

Since the diagonal elements are never zero we start by finding the extreme points which generate 2×2 positive definite submatrices of **B** in combination with an isolated third diagonal term. In this step let

$$\bar{\rho}_{12} = \bar{\rho}_{13} = \bar{\rho}_{23} = \frac{1}{2}$$

There are eight possible sequences for the diagonal terms. Consider one such sequence 16, 9, 1. Then six vertices may be associated with this sequence. They are

$$\begin{bmatrix} 16 & -6 & 0 \\ -6 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & 6 & 0 \\ 6 & 9 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & 0 & -2 \\ 0 & 9 & 0 \\ -2 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & 0 & 2 \\ 0 & 9 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 16 & 0 & 0 \\ 0 & 9 & -\frac{3}{2} \\ 0 & -\frac{3}{2} & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & 0 & 0 \\ 0 & 9 & \frac{3}{2} \\ 0 & \frac{3}{2} & 1 \end{bmatrix}$$

Repeating the above for the remaining seven diagonal sequences produces a total of 48 extreme points.

Each diagonal sequence also gives rise to eight 3×3 positive definite extreme points. Again considering the 16, 9, 1 sequence and choosing

$$\bar{\rho}_{12} = \bar{\rho}_{13} = \bar{\rho}_{23} = \frac{1}{3} \tag{3-83}$$

one obtains:

$$\begin{bmatrix} 16 & 4 & \frac{2}{3} \\ 4 & 9 & 1 \\ \frac{2}{3} & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & -4 & \frac{2}{3} \\ -4 & 9 & 1 \\ \frac{2}{3} & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & 4 & -\frac{2}{3} \\ 4 & 9 & 1 \\ -\frac{2}{3} & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & 4 & \frac{2}{3} \\ 4 & 9 & -1 \\ \frac{2}{3} & -1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 16 & 4 & -\frac{2}{3} \\ 4 & 9 & -1 \\ -\frac{2}{3} & -1 & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & -4 & \frac{2}{3} \\ -4 & 9 & -1 \\ \frac{2}{3} & -1 & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & 4 & -\frac{2}{3} \\ 4 & 9 & -1 \\ -\frac{2}{3} & -1 & 1 \end{bmatrix} \quad \begin{bmatrix} 16 & -4 & -\frac{2}{3} \\ -4 & 9 & -1 \\ -\frac{2}{3} & -1 & 1 \end{bmatrix}$$

Repeating this process for the other seven diagonal sequences yields 56 more extreme points for a grand total of 112.

Observe that an $n \times n$ matrix will always be positive definite if each off-diagonal term satisfies

$$a_{ij} \leq \frac{n-1}{n} \frac{\sqrt{a_{ii}a_{jj}}}{n-1} = \frac{\sqrt{a_{ii}a_{jj}}}{n} \qquad (3\text{-}84)$$

Thus as n becomes large the limiting degree of correlation allowed by (3-75) is $1/n$.

From example 3.3 it is clear that even the simplified technique presented here for generating convex subsets of V may still lead to an excessive number of extreme points. In any real problem much more may be known about Q and R than assumed here. Many elements may be known exactly. Others may be known to lie in very restricted ranges. In any event, the designer should restrict his selection of extreme points to the smallest set which adequately represents his uncertainty in the Q and R matrices.

CHAPTER IV

DETERMINATION OF THE MINIMAX FILTER FOR HIGHER ORDER SYSTEMS

In this chapter the computational problems associated with determining the infinite time minimax filter for higher order systems are discussed. A program for calculating the minimax filter gain is described and the error performance of several examples is presented.

4.1 Steepest Ascent Search for the Maximum

In Chapter II the infinite time optimal filter performance index $\bar{J}_0(v)$ and its first derivatives were shown to be continuous in the uncertain parameter set V and a linear algebraic equation for the first derivatives was derived. Steepest ascent techniques for the computational maximization of $\bar{J}_0(v)$ thus appear suitable.

It is well-known that a steepest ascent search finds only a local maximum. Some means must be provided to determine when a particular local maximum is also a global maximum.

We have already seen that the minimax value of $\bar{J}_M(K, v)$ occurs at a unique point (K_∞^*, v_∞^*) in $V_K \times V$ and that this in turn implies that $\max_{v \epsilon V} \bar{J}_0(v)$ occurs at the unique point v_∞^* in V. This uniqueness together with the fact that $\bar{J}_0(v)$ is concave in V are sufficient to guarantee that any local maximum is also a global maximum. This is shown as follows: Assume that $\bar{J}_0(v_1)$, $v_1 \neq v_\infty^*$ is a local maximum. Then

$$\bar{J}_0(v_\infty^*) > \bar{J}_0(v_1) \tag{4-1}$$

Now consider the line between v_1 and v_∞^* in V; i.e., the locus of points v_2 such that

$$v_2 = v_1 + \alpha(v_\infty^* - v_1); \quad 0 \leq \alpha \leq 1 \tag{4-2}$$

or

$$v_2 = \alpha v_\infty^* + (1-\alpha) v_1 \tag{4-3}$$

Then since V is convex, $v_2 \in V$. Because $\bar{J}_0(v)$ is continuous there must be a value of v_2 (and therefore α) in a neighborhood of v_1 such that

$$\bar{J}_0(v_2) < \bar{J}_0(v_1) < \bar{J}_0(v_1) + \alpha[\bar{J}_0(v_\infty^*) - \bar{J}_0(v_1)] \tag{4-4}$$

where the right-most inequality follows from (4-1). Regrouping the right-hand side of inequality (4-4) one obtains

$$\bar{J}_0(v_2) < \alpha \bar{J}_0(v_\infty^*) + (1-\alpha)\bar{J}_0(v_1) \tag{4-5}$$

But $\bar{J}_0(v)$ is concave in V requiring that

$$\bar{J}_0(v_2) \geq \alpha J_0(v_\infty^*) + (1-\alpha)\bar{J}_0(v_1) \tag{4-6}$$

Contradiction.

If any component of v_∞^* is an interior point of V then it is necessary that

$$\left.\frac{\partial \bar{J}_0(v)}{\partial v_i}\right|_{v = v_\infty^*} = 0 \tag{4-7}$$

By an argument similar to the above it can be shown that (4-7) cannot be zero at a distinct interior point $v \neq v_\infty^*$ without violating the concavity of $\bar{J}_0(v)$ in V. Thus $\bar{J}_0(v)$ is ideally suited to computational maximization via steepest ascent search.

4.2 A Steepest Ascent Program for Finding S_1

This section describes a first-order (gradient) steepest ascent program for finding max $\bar{J}_0(v)$.

A steepest-ascent or gradient search is an iterative algorithm for improving estimates of the maximizing parameter values. At each stage of the process an estimate of v_∞^*, say v_k, is used to compute $\bar{P}(v_k)$, $\bar{J}_0(v_k)$ and $\text{grad}_V \bar{J}_0(v)\big|_{v=v_k}$. This last quantity is used to obtain a revised estimate of v_∞^* denoted v_{k+1} where

$$v_{k+1} = v_k + \gamma \, \text{grad}_V \, \bar{J}_0(v)\Big|_{v=v_k} \tag{4-8}$$

and γ is some positive constant.

v_{k+1} is now substituted for v_k and the process is repeated until $\|\mathrm{grad}_V \bar{J}_o(v)\|$ becomes "small" or a boundary of V is attained. Computationally then the problem boils down to that of obtaining repeated solutions to the algebraic Riccati equation (2-57) for $\bar{P}(v_k)$ and the linear matrix equations (2-61) and (2-71) which yield $\mathrm{grad}_V \bar{J}_o$.

The equations to be solved at each iteration are repeated here for convenience.

$$0 = F\bar{P}_k + \bar{P}_k F^T + GQ_k G^T - \bar{P}_k H^T R_k^{-1} H\bar{P}_k \tag{4-9}$$

$$K_k = \bar{P}_k H^T R_k^{-1} \tag{4-10}$$

$$\bar{J}_o(k) = \mathrm{tr}\,[W\bar{P}_k] \tag{4-11}$$

$$0 = (F - K_k H)\bar{P}_{ij}(k) + \bar{P}_{ij}(k)(F - K_k H)^T + Q_{ij} \tag{4-12}$$

$$\left.\frac{\partial \bar{J}_o(k)}{\partial q_{ij}}\right|_{v=v_k} = \mathrm{tr}\,[W\bar{P}_{ij}] \tag{4-13}$$

$$0 = (F - K_k H)\bar{P}_{ij} + \bar{P}_{ij}(F - K_k H)^T + K_k R_{ij} K_k^T \tag{4-14}$$

$$\left.\frac{\partial J_o(v)}{\partial r_{ij}}\right|_{v=v_k} = \mathrm{tr}\,[W\bar{P}_{ij}(k)] \tag{4-15}$$

$$q_{ij}(k+1) = q_{ij}(k) + \gamma \left. \frac{\partial \bar{J}_o(v)}{\partial q_{ij}} \right|_{v=v_k} \tag{4-16}$$

$$r_{ij}(k+1) = r_{ij}(k) + \gamma \left. \frac{\partial \bar{J}_o(v)}{\partial r_{ij}} \right|_{v=v_k} \tag{4-17}$$

A Newton-Rhapsen iterative algorithm given by Blackburn [17] was used to solve (4-9). Since $\bar{P}(v_k)$ is available at each iteration and since $\|\bar{P}(v_{k+1}) - \bar{P}(v_k)\|$ is generally "small" this algorithm provided very rapid updating of the $\bar{P}$ matrix. Chen and Shieh [18] have presented an algorithm which converts the n^2-dimensional, linear matrix algebraic equation

$$A^T P + PA = -Q \tag{4-18}$$

into an equivalent $n(n+1)/2$-dimensional linear vector matrix equation. This algorithm was used in conjunction with a standard routine for solving sets of linear equations to compute solutions to (4-12) and (4-14).

A block diagram and a brief description of the maximizing program are given below. Referring to Fig. 4-1, input to MAXIMUM consists of the system F and H matrices, the weighting matrix W, the dimensions n and m, the initial values for Q, R, and $\bar{P}$, the scale factor γ, and the arrays QMAX and RMAX. The diagonal terms of Q and R are set to their maximum values. Only maximizing off-diagonal

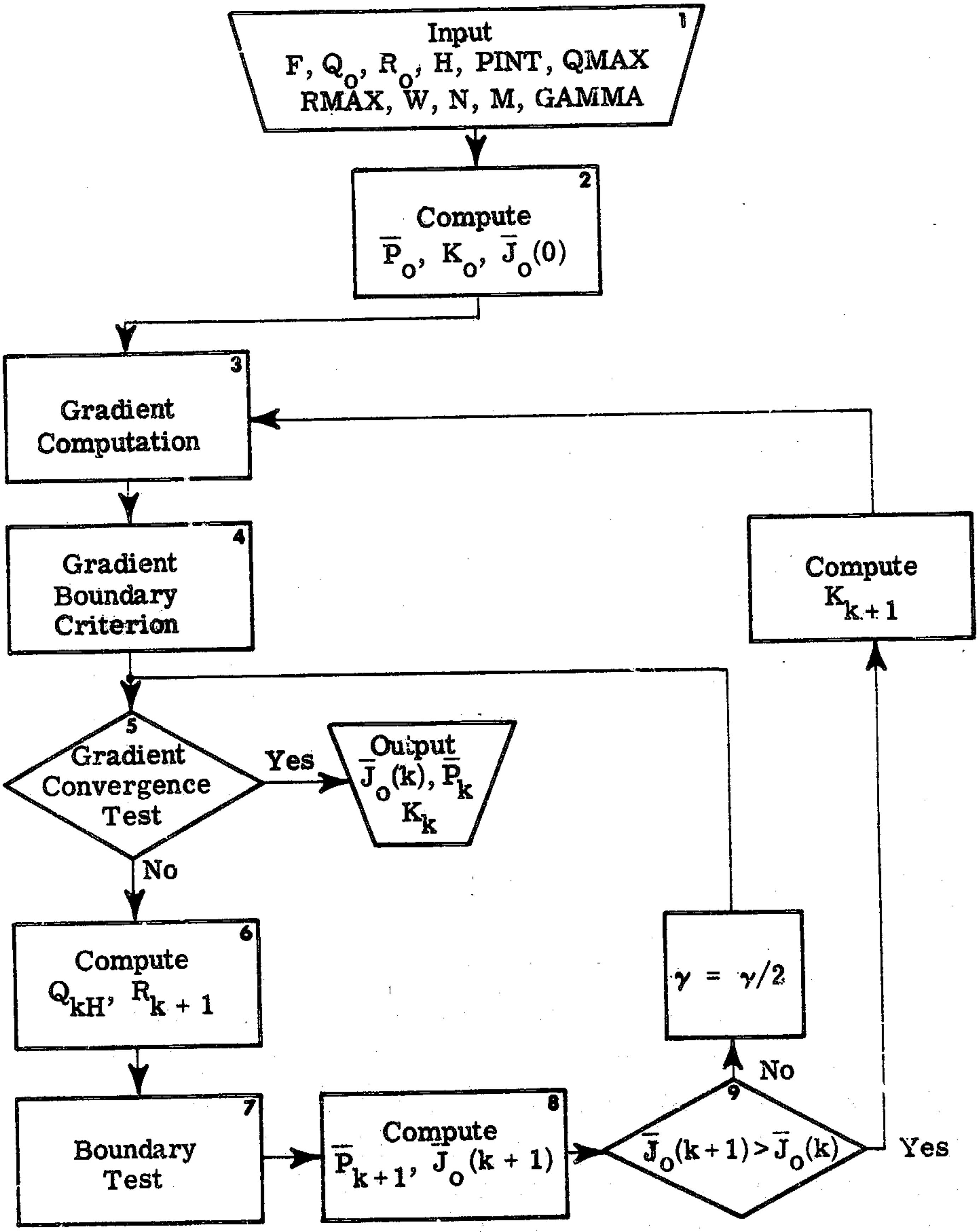


Figure 4-1 Program MAXIMUM

elements are determined by MAXIMUM. The arrays QMAX and RMAX contain the magnitude of the maximum allowed value of each off-diagonal element. For example,

$$\mathrm{QMAX}(I, J) = \frac{\bar{q}_{ij}}{N-1}$$

If a particular off-diagonal element of Q or R is known the corresponding element in QMAX or RMAX is set to zero. After the initial computation of $\bar{P}_o$, $\bar{K}_o$, and $\bar{J}_o$ using eqs. (4-9), (4-10), and (4-11) in block 2, the gradient of $\bar{J}_o$ is computed for each uncertain off-diagonal element of Q and R using (4-12) and (4-13) or (4-14) and (4-15). Skipping block 4 for the moment, a gradient convergence test is next executed in block 5. This test has the form:

$$\frac{\gamma \sum_{i=1}^{N} \left[\mathrm{grad}_{v_i} \bar{J}_o(v) \Big|_{v=v_k} \right]^2}{\bar{J}_o(v_k)} \leq \eta \qquad (4\text{-}19)$$

Thus if the estimated relative change in the performance index is less than η the present values of $\bar{P}$, $\bar{K}$, $\bar{J}_o$ are outputed and computation stops. If not, Q_k and R_k are updated to Q_{k+1} and R_{k+1} and a boundary test performed. In this test the various off-diagonal terms of Q_{k+1} and R_{k+1} are compared to their maximum allowed values. If the last increment has resulted in a particular term exceeding its allowed range it is reset to the nearest boundary. Elements on the boundary are tagged for later use in block 4. $\overline{P}_{k+1}$ and $\overline{J}_o(k+1)$ are then computed and a comparison between $\overline{J}_o(k)$ and $\overline{J}_o(k+1)$ is made in block 9. If $\overline{J}_o(k+1)$ is less than $\overline{J}_o(k)$ the scale factor γ is reduced and the loop is re-entered at block 5. If $\overline{J}_o(k+1)$ is greater than $\overline{J}_o(k)$ the trial is considered successful, K_{k+1} is computed, and a new trail begins by returning to block 3.

In block 4 the gradient with respect to each off-diagonal term on a boundary is tested. If the direction of increasing $\overline{J}_o$ is such as to carry a given term beyond its allowed bounds, the gradient of $\overline{J}_o$ with respect to that term is dropped from the gradient convergence test in block 5.

4.3 Some Numerical Examples

Program MAXIMUM was written in FORTRAN IV and run in a time-sharing mode on a CDC-3600 computer for systems up to order eight. MAXIMUM was used successfully to find the maximum for several higher order systems. Three examples are discussed in detail below:

Example 4.1

This example illustrates the ease with which the maximum can be found for a large number of uncertain variables. Consider a simple two integrator oscillator in which each integrator receives a random input composed of "white" noise plus a first-order Markov process.

Noisy measurements of the state of each integrator are available. The state and measurement equations are taken to be:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \tag{4-20}$$

and

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \tag{4-21}$$

The pairs (u_1, u_2), (u_1, u_3), (u_2, u_4), (u_3, u_4) and (w_1, w_2) are thought to be correlated. The diagonal elements of Q and R and the range of each off-diagonal element is taken to be

$$\{Q\} = \begin{bmatrix} 1 & \pm.30 & \pm.45 & 0 \\ \pm.30 & 1 & 0 & \pm.30 \\ \pm.45 & 0 & 2 & \pm.45 \\ 0 & \pm.30 & \pm.45 & 1 \end{bmatrix} \tag{4-22}$$

and

$$\left\{R\right\} = \begin{bmatrix} 1 & \pm 1 \\ \pm 1 & 2 \end{bmatrix} \tag{4-23}$$

where the range of each off-diagonal element was selected in accordance with the discussion in Section 3.3. The mean square error in estimating x_1 and x_2 was chosen for the performance index, i.e.

$$\overline{J}_o = E\left\{x_1^2 + x_2^2\right\} \tag{4-24}$$

With all off-diagonal terms initially set to zero and $\eta = 10^{-6}$, the program found the maximum in 24 iterations. Total running time: 29.5 seconds. The final results were:

$$Q_{max} = \begin{bmatrix} 1 & -.30 & .45 & 0 \\ -.30 & 1 & 0 & .30 \\ .45 & 0 & 2 & -.45 \\ 0 & .30 & -.45 & 1 \end{bmatrix}$$

$$\hat{R}_{max} = \begin{bmatrix} 1 & -.2664 \\ -.2664 & 2 \end{bmatrix}$$

$$K_\infty^* = \begin{bmatrix} .12769 & .08939 \\ .08603 & .92879 \\ .42533 & -.10866 \\ .03261 & .09618 \end{bmatrix}$$

$$\max \bar{J}_o = 3.08780$$

The next two examples are concerned with the error performance of terrestrial inertial navigators mechanized in local geographic coordinates. The indicated position, heading, and velocity of an inertial navigator are subject to random errors induced by random gyro drift rates, random accelerometer outputs, and random uncertainties in the local gravity vector, to name a few. Linear models relating input errors to the output errors are well known [19]. In an attempt to reduce the output errors external noisy measurements of position and velocity are taken. The Kalman filter is then used to estimate the state of the navigator error model and thus the errors in indicated output. Unfortunately, the statistics of the disturbing input and measurement noises are often not well known. Enter the minimax filter.

Example 4.2

In many instances the position error of an inertial navigator mechanized in local geographic coordinates is adequately described by a three-degree-of-freedom oscillator called the earth rate loop (ERL) model because it has a twenty-four hour period induced by the earth's rotation. The three states of this oscillator, ψ_x, ψ_y, and ψ_z represent the small angular misalignment of the platform about the true local coordinate system (see Fig. 4-2). Inputs the the ERL error model are

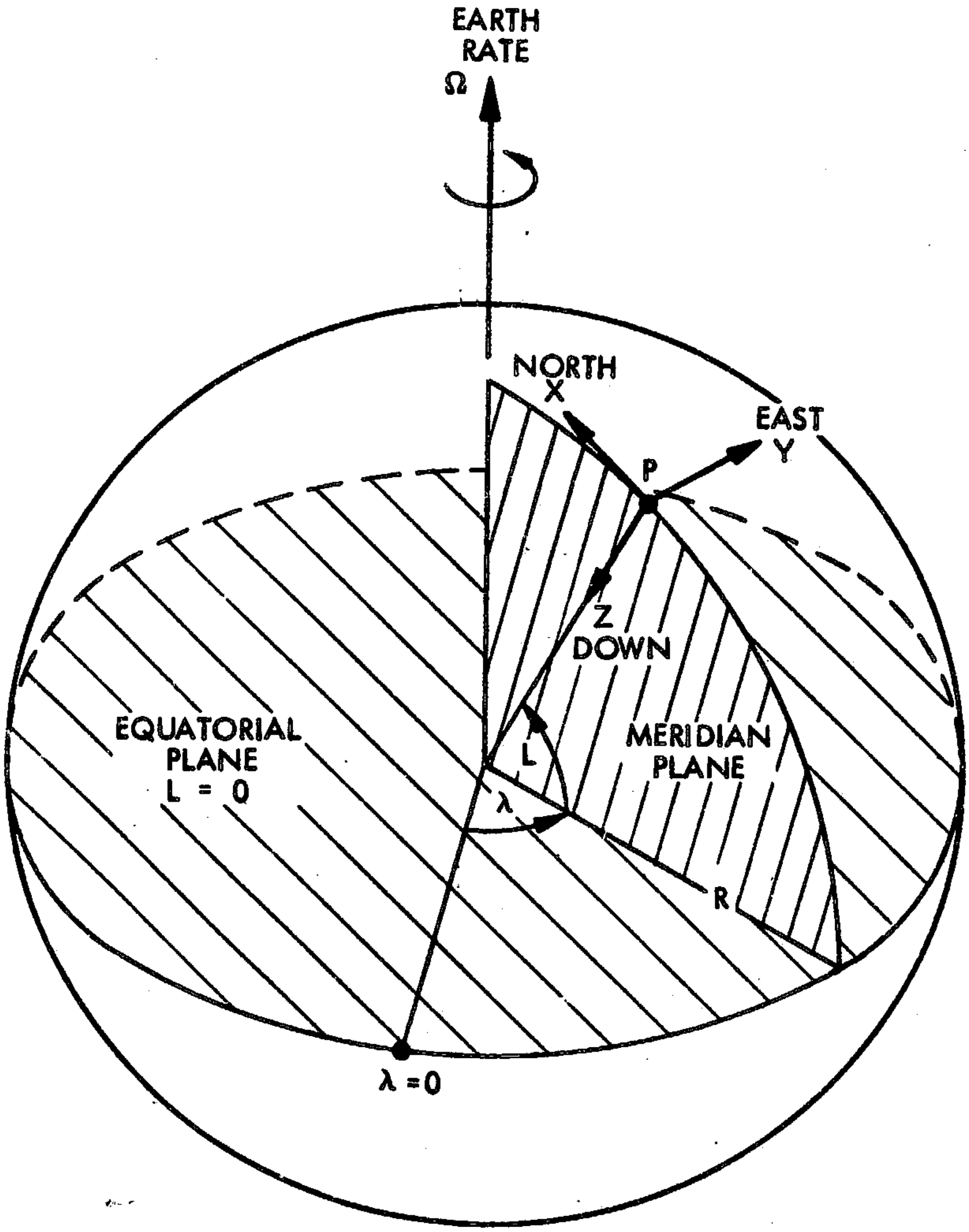


Figure 4-2 North-Vertical Coordinates

the x, y, and z gyro random drift rates which are assumed to be first-order continuous Markov processes. This process is generated in the model by a shaping filter appropriately excited by white noise. If ϵ represents a gyro drift rate then

$$\dot{\epsilon} = -\beta\epsilon + u \tag{4-25}$$

The variance of u is related to the steady-state variance of ϵ by

$$\sigma_u^2 = 2\beta\sigma_\epsilon^2 \tag{4-26}$$

The mean square values of the gyro drift rates are fairly well known. The x and y gyro drift rates are also known to be correlated, but the amount of correlation is uncertain.

Position error is related to the ψ_x and ψ_y tilt angles through the earth's radius, R_e as follows:

$$\delta X = R_e \psi_y$$

and

$$\delta Y = -R_e \psi_x \tag{4-27}$$

where δX and δY are the X and Y position errors. A figure of merit (FOM) for the navigator is its total rms radial position error

$$\delta R = R_e \left[E\left\{ \psi_x^2 + \psi_y^2 \right\} \right]^{\frac{1}{2}} \tag{4-28}$$

Noisy measurements of position are available. The rms magnitude of each measurement noise and the cross-correlation between them is uncertain.

A complete specification of the position error problem is given below:

Navigator Error State Equations

$$\begin{bmatrix} \dot{\psi}_x \\ \dot{\psi}_y \\ \dot{\psi}_z \\ \dot{\epsilon}_x \\ \dot{\epsilon}_y \\ \dot{\epsilon}_z \end{bmatrix} = \begin{bmatrix} 0 & -\Omega_V & 0 & K_1 & 0 & 0 \\ \Omega_V & 0 & \Omega_H & 0 & K_1 & 0 \\ 0 & -\Omega_H & 0 & 0 & 0 & K_1 \\ 0 & 0 & 0 & -\beta_x & 0 & 0 \\ 0 & 0 & 0 & 0 & -\beta_y & 0 \\ 0 & 0 & 0 & 0 & 0 & -\beta_z \end{bmatrix} \begin{bmatrix} \psi_x \\ \psi_y \\ \psi_z \\ \epsilon_x \\ \epsilon_y \\ \epsilon_z \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

Measurement Equations

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0 & R_e & 0 & 0 & 0 & 0 \\ -R_e & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \psi_x \\ \psi_y \\ \psi_z \\ \epsilon_x \\ \epsilon_y \\ \epsilon_z \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

where

ψ_x, ψ_y, ψ_z = platform misalignment (rad)

$\epsilon_x, \epsilon_y, \epsilon_z$ = gyro drift rates (deg/hr)

Ω = earth rate = .262 rad/hr

$\Omega_V = \Omega \sin L$

$\Omega_H = \Omega \cos L$

L = lattitude (deg)

K_1 = .01734 rad/deg

R_e = 3437 nautical miles (nm)

$\sigma_{\epsilon_x} = \sigma_{\epsilon_y} = .0014$ deg/hr

$\beta_x = \beta_y = 1/\text{hr}$

$\rho_{\epsilon_x \epsilon_y} \le .75$

$\sigma_{\epsilon_z} = .003$ deg/hr

$\beta_z = .1/\text{hr}$

$.1 \le \sigma_{w_1} \le .25$ nm

$.1 \le \sigma_{w_2} \le .25$ nm

$\rho_{w_1 w_2} \le .8$

The diagonal elements of Q and R and the range of the uncertain off-diagonal elements are computed using eq. (4-26). The sets V_Q and V_R for this problem are:

$$V_Q = \begin{bmatrix} .4 \times 10^{-5} & \pm.3 \times 10^{-5} & 0 \\ \pm.3 \times 10^{-5} & .4 \times 10^{-5} & 0 \\ 0 & 0 & .18 \times 10^{-5} \end{bmatrix}$$

$$V_R = \begin{bmatrix} .0625 & \pm.05 \\ \pm.05 & .0625 \end{bmatrix}$$

where the maximal values are utilized for the diagonal elements of V_R. A minimax radial error of .269 nautical miles was attained at the point

$$q_{12} = .3 \times 10^{-6} \tag{4-29a}$$

$$r_{12} = .0234 \tag{4-29b}$$

The S_2 filter was also found for this example using ad hoc procedure. A minimax absolute radial deviation from optimum of .026 nautical miles was attained. The design point for this filter is

$$q_{12} = -.15 \times 10^{-6}$$

$$r_{12} = -.03$$

where it was assumed that the diagonal elements of V_R were again maximum.

The error performance of both filters is shown in Table 4.1. The "+" and "-" signs in the first column of this table refer to the extreme values of the unknown off-diagonal terms. Notice that the S_1 filter error is insensitive to the value of the off-diagonal terms since it is designed for the point where the gradient with respect to these terms is zero.

The lower half of Table 4.1 summarizes the S_1, S_2 comparison. The S_1 filter provides a least upper bound or radial position error at the expense of greater maximum deviation from optimality. The S_2 filter to the contrary minimizes the maximum deviation from optimality at the expense of greater absolute error. System specifications would dictate which of the two filters would be preferable in a given application.

Table 4.1

Position Error FOM for Example 4.2
(nautical miles)

UNCERTAIN PARAMETER q_{12}, r_{12}	OPTIMAL FILTER	S_1 FILTER	S_2 FILTER	DEVIATION FROM OPTIMUM Δ_1	Δ_2
+, +	.258	.269	.284	.011	.026
+, -	.221	.269	.247	.238	.026
-, +	.262	.269	.283	.007	.021
-, -	.220	.269	.246	.039	.026
origin	.265	.269	.265	.004	~0
Max Error	.269	.269	.284		
Max Δ	--	.039	.026		

Table 4.2 shows the error performance of the S_1 filter when the rms value of the position measurement noise is reduced to .1 nautical miles in both x and y directions. The maximum optimum error was found for this condition to occur at

$$q_{12} = -.509 \times 10^{-6}$$

$$r_{12} = .00213$$

This point is listed as "max opt" in Table 4.2. Observe that while the optimum FOM ranges from .125 to .269 nautical miles, the minimax filter FOM never exceeds optimum by more than .05 nautical miles.

Table 4.2

Position Error FOM for Example 4.2 with Reduced Measurement Noise (nautical miles)

UNCERTAIN PARAMETER q_{12}, r_{12}	OPTIMAL FILTER	S_1 FILTER	DEVIATION
+, +	.133	.175	.042
+, -	.131	.175	.044
-, +	.140	.175	.035
-, -	.125	.175	.050
origin	.144	.175	.031
Max Opt	.144	.175	.031

These results are valid for 45° latitude only; however, they change rather slowly with latitude. In actual practice the minimax filter could be found for latitude increments of, say 5°. One would then use S_1 filter designed for a latitude closest to his indicated latitude.

Example 4.3:

Velocity error propagation in an inertial navigator is largely governed by the so-called "Schuler Loop" dynamics [19]. The process of forcing an inertial platform to remain perpendicular to the local vertical gives rise to Schuler oscillations. Since perpendicularity is maintained by controlling alignment about both the x and y axes there are actually two Schuler oscillators which are cross-coupled through inertial angular rates about the local vertical. Random gyro drift rates, accelerometer noise, and random angular misalignment between the local geographic vertical and the local gravity vector (deflection of the vertical) constitute the major error inputs to the Schuler loops. All of the above errors are modeled as continuous first-order Markov random processes. As before, the x and y gyro drift rates are known to be correlated. Deflection of the vertical is a two-dimensional spatial random process with uncertain rms values and uncertain cross-correlation. In the error analysis this spatial random process is converted into a time random process by assuming constant vehicle speed.

In order to reduce velocity errors and damp the Schuler oscillations measurements of vehicle velocity over the earth's surface are taken. On a surface ship what is actually measured is the ship's velocity relative to the water. This measurement differs from true velocity by the local ocean currents. Ocean currents, like gravity anomalies are modeled as two-dimensional spatial random processes with uncertain rms values and cross-correlations. The performance index of interest in this example is the total radial velocity error

$$\delta V_R = \left[E\left\{\delta V_x^2 + \delta V_y^2\right\}\right]^{\frac{1}{2}} \tag{4-30}$$

A complete specification of this example is given below:

Schuler Loop Error State Equations

$$
\begin{bmatrix} \delta\dot{\theta}_x \\ \delta\dot{V}_x \\ \delta\dot{\theta}_y \\ \delta\dot{V}_y \\ \dot{\epsilon}_x \\ \dot{\epsilon}_y \\ \dot{\epsilon}_{ax} \\ \dot{\epsilon}_{ay} \end{bmatrix}
=
\begin{bmatrix}
0 & \frac{1}{R_e} & 0 & 0 & K_1 & 0 & 0 & 0 \\
-g & 0 & 0 & -2\Omega_v & 0 & 0 & K_2 & 0 \\
0 & 0 & 0 & \frac{1}{R_e} & 0 & K_1 & 0 & 0 \\
0 & 2\Omega_v & -g & 0 & 0 & 0 & 0 & K_2 \\
0 & 0 & 0 & 0 & -\beta_x & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -\beta_y & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -\beta_a & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -\beta_a
\end{bmatrix}
\begin{bmatrix} \delta\theta_x \\ \delta V_x \\ \delta\theta_y \\ \delta V_y \\ \epsilon_x \\ \epsilon_y \\ \epsilon_{ax} \\ \epsilon_{ay} \end{bmatrix}
+
\begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}
$$

Velocity Error Measurement Equations

$$
\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}
=
\begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix} \delta\theta_x \\ \delta V_x \\ \delta\theta_y \\ \delta V_y \\ \epsilon_x \\ \epsilon_y \\ \epsilon_{ax} \\ \epsilon_{ay} \end{bmatrix}
+
\begin{bmatrix} \epsilon_{vx} \\ \epsilon_{vy} \end{bmatrix}
$$

Measurement Error States

$$\dot{\epsilon}_{vx} = -\beta_v \epsilon_{vx} + w_{vx}$$

$$\dot{\epsilon}_{vy} = -\beta_v \epsilon_{vy} + w_{vy}$$

where

$\delta\theta_x, \delta\theta_y$ = computed tilt angles (rad)

$\delta V_x, \delta V_y$ = velocity error (knots)

ϵ_x, ϵ_y = gyro drift rates (deg/hr)

$\epsilon_{ax}, \epsilon_{ay}$ = vertical deflection errors ($\widehat{\sec}$)

$\epsilon_{vx}, \epsilon_{vy}$ = ocean current errors (knots)

g = gravity (68,635 knots/hr)

$K_2 = 3.329$ knots/sec/$\widehat{\sec}$

$\beta_x = \beta_y = 1/\text{hr}$

$\beta_a = .2/\text{hr}$

$\beta_v = .2/\text{hr}$

$\sigma_{\epsilon_x} = \sigma_{\epsilon_y} = .001$ deg/hr

$\rho_{\epsilon_x \epsilon_y} \le .75$

$5 \le \sigma_{\epsilon_{ax}}, \ \sigma_{\epsilon_{ay}} \le 10$ $\widehat{\sec}$

$\rho_{\epsilon_{ax}\epsilon_{ay}} \le .75$

$1 \le \sigma_{\epsilon_{vx}}, \ \sigma_{\epsilon_{vy}} \le 2$ knots

$\rho_{\epsilon_{vx}\epsilon_{vy}} \le .75$

The measurement noises in this example are "colored." Following the technique of Bryson and Johanson [20] auxiliary measurements are generated which contain only the original system states plus white noise. These auxiliary measurements are:

$$z_1' = \dot{z}_1 + \beta_v z_1 = -g\delta\theta_x - 2\Omega_v \delta V_y + K_2 \epsilon_{ax} + \beta_v \delta V_x + w_{vx}$$

$$z_2' = \dot{z}_2 + \beta_v z_2 = -g\delta\theta_y + 2\Omega_v \delta V_x + K_2 \epsilon_{ay} + \beta_v \delta V_y + w_{vy} \tag{4-31}$$

or

$$\begin{bmatrix} z_1' \\ z_2' \end{bmatrix} = \begin{bmatrix} -g & \beta_v & 0 & -2\Omega_v & 0 & 0 & K_2 & 0 \\ 0 & 2\Omega_v & -g & \beta_v & 0 & 0 & 0 & K_2 \end{bmatrix} \begin{bmatrix} \delta\theta_x \\ \vdots \\ \epsilon_{ay} \end{bmatrix} + \begin{bmatrix} w_{vx} \\ w_{vy} \end{bmatrix} \tag{4-32}$$

The problem is analyzed using the original state equations with the derived measurements of eq. (4-32). The sets V_Q and V_R for this example are:

$$V_Q = \begin{bmatrix} .2\times10^{-5} & \pm.15\times10^{-5} & 0 & 0 \\ \pm.15\times10^{-5} & .2\times10^{-5} & 0 & 0 \\ 0 & 0 & 40 & \pm30 \\ 0 & 0 & \pm30 & 40 \end{bmatrix}$$

and

$$V_R = \begin{bmatrix} 1.6 & \pm1.2 \\ \pm1.2 & 1.6 \end{bmatrix}$$

In this example the minimax occurs when all uncertain cross-correlations are zero. In this respect the example is similar to example 2.3 and the comments given there apply in this instance. Tables 4.3 and 4.4 list the velocity error performance of the S_1 filter when the rms vertical deflection and ocean current errors are at their maximum and minimum values respectively. The tabulated results are fairly

Table 4.3

Velocity Error FOM for Example 4.3
(knots)

UNCERTAIN PARAMETERS q_{12}, q_{34}, r_{12}	OPTIMAL FILTER	S_1 FILTER	DEVIATION
origin	.4963	.4963	–
−, −, −	.4894	.4963	.0069
−, −, +	.3964	.4963	.0999
−, +, −	.4063	.4963	.0900
−, +, +	.4895	.4963	.0068
+, −, −	.4895	.4963	.0068
+, −, +	.4063	.4963	.0900
+, +, −	.3964	.4963	.0999
+, +, +	.4894	.4963	.0069

Table 4.4

Velocity Error FOM for Example 4.3
With Reduced Noise
(knots)

UNCERTAIN PARAMETERS q_{12}, q_{34}, r_{12}	OPTIMAL FILTER	S_1 FILTER	DEVIATION
origin	.2602	.2605	.0003
−, −, −	.2570	.2605	.0033
−, −, +	.2079	.2605	.0526
−, +, −	.2445	.2605	.0360
−, +, +	.2555	.2605	.0050
+, −, −	.2555	.2605	.0050
+, −, +	.2245	.2605	.0360
+, +, −	.2079	.2605	.0526
+, +, +	.2570	.2605	.0033

self-evident. Observe that in all instances the minimax filter error is within 25% of the optimum error.

In summary, the utility of steepest ascent techniques for finding S_1 filters has been established. A detailed sensitivity analysis was presented for two significant higher-order examples. In each case the S_1 filter gives a least upper bound on system performance while providing near optimum performance over wide parameter variations.

CHAPTER V

ADDITIONAL TOPICS IN MINIMAX FILTERING

In this chapter some additional topics in minimax filtering are discussed for which only partial results are available. As such, these results represent points of departure for future research.

5.1 The Minimax Filter for Time Varying Statistics

In this section the minimax filtering problem for time varying Q and R is presented. The time variation may be arbitrary or uncertain. The primary tool in the analysis to be presented is the maximum principle of Pontryagin [21].

The plant, measurement, and filter equations are identical to those given in Chapter II. F, G, and H may now, however, have known time variation. The estimation error covariance still satisfies eq. (2-15) which is repeated here for convenience.

$$\dot{M} = (F - KH)M + M(F - KH)^T + GQG^T + KRK^T; \quad M(t_o) = P_o \tag{5-1}$$

The sets V_Q, V_R, and V first defined in Chapter II are redefined as follows:

$$V_Q(t) = \Big\{Q(t)\,|\,Q = Q^T, \quad Q \geq 0, \text{ and } 0 \leq q_{min}(t) \leq \text{tr}\, Q(t) \leq q_{max}(t) < \infty\Big\} \tag{5-2}$$

$$V_R(t) = \Big\{R(t) \,|\, R = R^T, \quad R > 0, \quad \text{and} \quad 0 < r_{min}(t) \le \operatorname{tr} R(t) \le r_{max}(t) < \infty\Big\} \tag{5-3}$$

and

$$V(t) = V_Q(t) \times V_R(t) \tag{5-4}$$

The filter performance index is taken to be

$$J_M(T) = \operatorname{tr}[W(T)\,M(T)]; \quad W(T) > 0 \tag{5-5}$$

where W(T) is now allowed to be time varying. The minimax performance index filter for time varying noise statistics may now be defined.

Def. 5.1: The minimax performance index filter is specified by that time trajectory in V_K, denoted by $[K^*(t)]$ for which

$$\max_{[v(t)]\epsilon V(t)} J_M\Big([K^*(t)], [v(t)], P_o, t_o, T\Big) = \min_{[K(t)]\epsilon V_K} \; \max_{[v(t)]\epsilon V(t)} J_M\Big([K(t)], [v(t)], P_o, t_o, T\Big) \tag{5-6}$$

The brackets around v(t) in (5-6) emphasize that the maximizing solution for v(t) is not a fixed value of v as in Chapter II, but a time history (i.e., trajectory) in V(t) running from t_o to T. This maximizing trajectory will be denoted by $[v^*(t)]$.

The minimaximization of $J_M(T)$ may be accomplished, at least in theory, via the maximum principle of Pontryagin. Consider the elements of the covariance matrix, M(t), as state variables in this problem and define a Hamiltonian as follows:

$$H = \sum_{i=1}^{n} \sum_{j=1}^{n} \dot{m}_{ij}\lambda_{ij} = \mathrm{tr}\left[\dot{M}\Lambda^T\right] = \mathrm{tr}\left[\Lambda\dot{M}\right] \tag{5-7}$$

Then using (5-1) H becomes

$$H(K,v,t) = \mathrm{tr}\left[\Lambda(F - KH)M + \Lambda M(F - KH)^T + \Lambda GQG^T + \Lambda KRK^T\right] \tag{5-8}$$

Necessary conditions for the minimaximizing trajectories are now found by minimaximizing H in V_K and $\dot{V}(t)$ for every t.

According to the maximum principle the costate matrix must satisfy the linear matrix differential equation

$$\dot{\Lambda} = -\left.\frac{\partial H}{\partial M}\right|_{K^*,v^*} = -\Lambda(F - K^*H) - (F - K^*H)^T\Lambda \tag{5-9}$$

with the transversality condition

$$\Lambda(T) = \left.\frac{\partial J_M}{\partial M}\right|_{t=T} = W(T) \tag{5-10}$$

Let $\Psi(t, t_o)$ denote the state transition matrix satisfying

$$\dot{\Psi}(t, t_o) = -(F - K^*H)^T\Psi(t, t_o); \quad \Psi(t_o, t_o) = I \tag{5-11}$$

then

$$\Lambda(T) = \Psi(T, t)\Lambda(t)\Psi^T(T, t); \quad t < T \tag{5-12}$$

Solving for $\Lambda(t)$ and using (5-10) one obtains

$$\Lambda(t) = \Psi(t, T)W(T)\Psi^T(t, T); \quad t < T \tag{5-13}$$

Now since $W(T) > 0$, and $\Psi(t, T)$ is never singular [14], $\Lambda(t)$ will also be positive definite.

Comparing (5-8) with (2-30) one sees that the expression for H is identical to that for $\dot{J}_M$ in Chapter II with W replaced by $\Lambda(t)$. It follows from the argument developed in Chapter II for $\dot{J}_M$ that H is strictly convex in V_K. Since H is linear in Q and R it is concave in V. Thus H satisfies the sufficient conditions of theorem 2.1 and

$$\min_{K\epsilon V_K} \max_{v\epsilon V(t)} H(K, v, t) = \max_{v\epsilon V(t)} \min_{K\epsilon V_K} H(K, v, t) \tag{5-14}$$

Performing the minimization of H over V_K first one has

$$\frac{\partial H}{\partial K} = -2\Lambda MH^T + 2\Lambda KR = 0 \tag{5-15}$$

Since Λ is positive definite, satisfying (5-15) requires that

$$MH^T = KR$$

or

$$K = MH^T R^{-1} \tag{5-16}$$

Substituting (5-16) into (5-1) produces

$$\dot{M} = MF + FM^T + GQG^T - MH^TR^{-1}HM; \quad M(t_o) = P_o \tag{5-17}$$

Now comparing (5-17) with (1-9) shows that

$$M(t) \equiv P(t); \quad t_o \le t \le T \tag{5-18}$$

and therefore $K^*(t) = K_o(t)$. Thus the problem reduces to that of maximizing the optimal filter performance index over the set of all trajectories $[V(t)]$. Although similar in nature, this result should not be confused with that obtained in Chapter II where v was assumed constant.

Using (5-18), H can be written as

$$H_o = \mathrm{tr}\,[\Lambda FP + \Lambda PF^T + \Lambda GQG^T - \Lambda PH^TR^{-1}HP] \tag{5-19}$$

where we wish to find

$$H_o(v^*, t) = \max_{v \epsilon V(t)} H_o(v, t) \tag{5-20}$$

The partial derivative of H_o with respect to an element of R is

$$\frac{\partial H_o}{\partial r_{ij}} = \mathrm{tr}\,[\Lambda PH^TR^{-1}R_{ij}R^{-1}HP] \tag{5-21}$$

where R_{ij} is defined in eq. (2-67). When i = j, R_{ii} is positive semi-definite and therefore so is $PH^TR^{-1}R_{ij}R^{-1}HP$. Λ is always positive definite. It can be shown that if $A \geq 0$ and $B > 0$ then

$$\mathrm{tr}\,[AB] > 0 \tag{5-22}$$

Thus for $i = j$, eq. (5-21) is always positive and the maximizing value of $r_{ii}(t)$ is simply its largest admissible value at time t.

A similar result for the diagonal terms of Q is obtained in exactly the same way. However, since H_o is linear in Q, the maximizing values of q_{ij} will in general be "bang-bang."[1] For example, let G = I. Then

$$\mathrm{tr}\,[\Lambda Q] = \sum_{i,j}^{n} q_{ij}\lambda_{ij} \tag{5-23}$$

and assuming $\lambda_{ij}(t) \not\equiv 0$, for a finite interval $(t_1, t_2) \subset (t_o, T)$, one has

$$q^*_{ij}(t) = \begin{Bmatrix} [q_{ij}(t)]_{max}; & \lambda_{ij}(t) > 0 \\ [q_{ij}(t)]_{min}; & \lambda_{ij}(t) < 0 \end{Bmatrix} \tag{5-24}$$

Furthermore, since $\lambda_{ii}(t) > 0$

$$q^*_{ii}(t) = [q_{ii}(t)]_{max} \tag{5-25}$$

Since H_o is not an explicit function of time it must be constant along the minimaximizing trajectory. This constant can be shown to be zero when $T \to \infty$ if the plant is uniformly observable and controllable.

---

[1] Singular problems can arise, however. See example 2.3 of Chapter II.

Let $\Phi_K(t, t_o)$ be the state transition matrix for the Kalman filter, i.e.,

$$\dot{\Phi}_K(t, t_o) = (F - K_oH)\Phi_K(t, t_o); \quad \Phi(t_o, t_o) = I \tag{5-26}$$

Then

$$\Psi(t, T) = \Phi^T(T, t)$$

so that (5-13) becomes

$$\Lambda(t) = \Phi^T(T, t)\, W(T)\, \Phi(T, t) \tag{5-27}$$

Using the matrix norm defined in Appendix A, one has

$$\|\Lambda(t)\| \leq \|W(T)\| \cdot \|\Phi_K(T, t)\|^2 \tag{5-28}$$

and

$$\|\Lambda(t)\| \leq \lim_{\substack{T\to\infty \\ t<\infty}} \|W(T)\| \cdot \|\Phi_K(T, t)\|^2 \tag{5-29}$$

Under the above assumptions the Kalman filter is uniformly assymptotically stable. Thus the limit on the left is zero which implies that $\Lambda(t) = 0$, $t < \infty$. This in turn requires that

$$H_o = \mathrm{tr}\,[\dot{P}\Lambda(t)] = 0; \quad t < \infty \tag{5-30}$$

so that H_o must be zero for all t.

In summary then, application of the maximum principle to the minimax filtering problem for time varying statistics shows that the minimax filter is found by maximizing the optimal filter performance index over the set of admissible trajectories in the uncertain parameter space. Application of the maximum principle leads to the two-point boundary value problem (TPBP) given by eqs. (5-9) and (5-17). Solution of this TPBP requires a complete specification of the maximizing control law. So far, it has been shown that this law is usually "bang-bang" in V_Q and that the diagonal elements of Q(t) and R(t) must be set to their maximum admissible values at each point in time.

If the plant is time invariant, and R and Q are purely diagonal, and the bounds defining V are constant, the infinite time minimax filter gain is also constant. Thus for large T one can replace the time varying optimal filter with a constant filter which places a least upper bound on the estimation error regardless of the exact time variation of Q and R. This is a useful, but presently limited result. The validity of this result for time varying Q and R matrices with non-zero off-diagonal elements should be investigated.

5.2 The Minimax Filter for Uncertain Dynamics

The theory of minimax filtering for large uncertainties in plant dynamics is largely untouched. A simple example, presented by the author in reference 22, however, suggests that the minimax approach will be quite fruitful in the presence of this type of uncertainty. The example is discussed in more detail below. Although minmax equals maxmin for this example, it is shown that the filter performance index does not meet the sufficient conditions of theorem 2.1.

Consider the first-order linear time invariant plant

$$\dot{x}(t) = -\rho x(t) + u(t) \tag{5-31}$$

with noisy measurement

$$z(t) = x(t) + w(t) \tag{5-32}$$

where u and w are zero mean Gaussian white noise processes for which $\mathrm{cov}(u) = q\delta(t-\tau)$ and $\mathrm{cov}(w) = r\,\delta(t-\tau)$ and the expected value of x(0) is zero. The parameter ρ is constant but uncertain. It is assumed to lie in the set

$$S_\rho = \left\{\rho \,|\, 0 < \rho_{min} \le \rho \le \rho_{max} < \infty\right\} \tag{5-33}$$

The filter chosen to estimate x(t) then has the form

$$\dot{\overline{x}}(t) = -\rho_f \overline{x}(t) + k_f[z(t) - \overline{x}(t)] \tag{5-34}$$

Let the parameter set (ρ_f, k_f) be represented by the vector β. For the moment we require only that the filter (5-34) be stable ($\rho_f + k_f > 0$) and that β lie in a compact set S_β of sufficient size to generate the Kalman filter for every $\rho \epsilon S_\rho$.

Let the steady-state covariance of $x(t) - \bar{x}(t)$ be $m(\rho, \beta)$.[2]

Then it is shown in the reference that

$$\min_{\beta \epsilon S_\beta} \max_{\rho \epsilon S_\rho} m(\rho, \beta) = \max_{\rho \epsilon S_\rho} \min_{\beta \epsilon S_\beta} m(\rho, \beta) = \max_{\rho \epsilon S_\rho} \bar{p}_0(\rho) \qquad (5\text{-}35)$$

However, differentiating eq. 8 of [22] with respect to ρ one obtains

$$\frac{\partial^2 m}{\partial \rho^2} = \frac{2\rho_f^2\left[\left(\rho_f + k_f\right)^2 + 3\rho\left(\rho + \rho_f + k_f\right)\right]}{\left[\rho\left(\rho + \rho_f + k_f\right)\right]^3} > 0 \qquad (5\text{-}36)$$

Thus $m(\rho, \beta)$ is convex in S_ρ which violates the sufficient conditions of theorem 2.1. The significance of this example, which is essentially negative, is this: If a generalization of (5-35) exists for plants of arbitrary order with uncertain F matrices, that result cannot be demonstrated by an appeal to theorem 2.1 of Chapter II.

[2] Since $\rho \neq \rho_f$, eq. (2-15) for $\mathrm{cov}\,(x - \bar{x})$ is no longer valid. In general, for systems of order n one must solve the linear variance equation associated with the 2n order system consisting of the coupled plant and filter state equations.

CHAPTER VI

SUMMARY AND SUGGESTIONS

6.1 Summary

Minimax criteria for the design of low sensitivity linear filters for state estimation in the presence of large parameter uncertainties have been proposed. The case of constant but uncertain plant and measurement noise statistics has been fully explored. With regard to the S_1 criterion it was shown that the value of the minimax filter return function is equal to the maximum value of the optimal filter return function over the uncertain parameter set. Furthermore, the minimax filter was shown to be the Kalman filter for the maximizing set of parameters. Certain properties of the infinite time maximization problem were then developed. It was established that the optimal return function is continuous in the uncertain parameter set and the gradient of this return function with respect to the uncertain parameters was found to always exist. Using the uniqueness of the minimax point together with the continuity and concavity of $\bar{J}_o(v)$ in V, the optimal return function was shown to possess only a global maximum. A straight forward steepest ascent search was then used to find minimax filters for several higher order examples.

Again for constant but undertain noise statistics, the infinite time minimax sensitivity filters (S_2 and S_3 criteria) were shown to be unique and the minimax sensitivity filter gain was found to be optimal for at least one value of the uncertain parameters. Unfortunately, min-max does not equal max-min for the S_2 and S_3 filters and one is thus forced to solve the complete min-max problem. It was shown, however, that the maximum of S_2 and S_3 is attained over a finite set of points in V called the extreme points of V, thereby greatly reducing the search problem in that domain. The minimax filtering problem for plant and measurement noise with uncertain or arbitrary time variation was investigated in some detail. It was determined that the minimax performance index can be found by maximizing the optimal filter performance index over the set of admissable trajectories in the uncertain parameter space. Furthermore, to maximize the optimal filter performance index, the diagonal elements of Q(t) and R(t) must assume their maximum admissable values at each point in time. Finally, if the plant dynamics are time-invariant, Q(t) and R(t) are purely diagonal, and the infinite time bounds on V are constant, the infinite time minimax filter gain for time varying Q and R was shown to be constant.

6.2 Suggestions for Further Research

Given the present level of understanding three areas in continuous time minimax filtering appear most likely to yield worthwhile results. At this point no comprehensive study of the computational difficulties associated with finding minimax sensitivity filters of Chapter III has been made. Salmon [10] has given a new algorithm for solving algebraic minimax problems without a saddle point and proved its convergence. Taking advantage of the special properties of the minimax sensitivity filters may result in some simplification of this algorithm. A major effort, however, is still required to translate this algorithm into a working design technique for systems of arbitrary order.

As indicated in section 5.1, under certain fairly stringent conditions on the form of Q(t) and R(t) one can, for large T, replace the time varying optimal filter with a constant filter which places a least upper bound on the performance index regardless of the exact time variation of Q or R. This is a useful, but presently limited result. Efforts should be made to determine whether or not this result can be extended to more general forms of Q and R.

The theory of minimax filtering in the presence of large uncertainties in plant dynamics is almost wholly untouched. The simple example in reference [22] suggests, however, that minimax design criteria are quite appropriate under this type of uncertainty. The theoretical and computational aspects of minimax filter design for plant dynamic uncertainties should be explored.

Finally, this dissertation has dealt exclusively with continuous time systems and continuous time filters. It seems clear that appropriate discrete time analogs must exist for most, if not all, of the results obtained in Chapters II and III. The minimax filtering problem should be formulated for discrete time systems and the existence of these analogous results determined.

APPENDIX A

A PROPERTY OF V_Q AND V_R

A.1 Introduction [23, 24]

Let $\mathcal{M}$ be the set of all real $n \times n$ matrices. Then $\mathcal{M}$ is identical with n^2-dimensional Euclidean space. An inner product may be defined on $\mathcal{M}$ as follows [23]:

Def. A.1: Let A, $B \in \mathcal{M}$. Then the inner product of A with B is

$$\langle A, B\rangle = \operatorname{tr}(AB^T) \tag{A-1}$$

This inner product induces a norm

$$\|A\| = \sqrt{\operatorname{tr}(AA^T)} \tag{A-2}$$

and a metric

$$d(A, B) = \|A - B\| \tag{A-3}$$

The pair $(\mathcal{M}, d)$ forms a complete metric space with the metric topology induced by $d(\cdot)$.

We shall have need of a second norm on $\mathcal{M}$, often called the Sup norm, which is defined as

$$\|A\|_\infty = \sup_{\|x\|_E \leq 1} \|Ax\|_E; \quad x \in E_n \tag{A-4}$$

where $\|\cdot\|_E$ denotes the conventional Euclidean norm in E_n. It is known that norms (A-2) and (A-4) are topologically equivalent. For our purposes this means that convergence in one implies convergence in the other. Fundamental norm properties required in the next section are listed below:

$$\|\lambda A\| = |\lambda|\,\|A\|; \quad \lambda \epsilon E_1$$

$$\|AB\| \leq \|A\| \cdot \|B\| \tag{A-5}$$

$$\|A+B\| \leq \|A\| + \|B\|$$

A.2 A Theorem

Let $\mathbf{A}$ be the set of all real symmetric positive-definite matrices such that

$$0 < b \leq tr(A) \leq c < \infty; \quad \forall A \epsilon \mathbf{A} \tag{A-6}$$

Then $\mathbf{A}$ is bounded and closed.

(1) Proof $\mathbf{A}$ is bounded:

Using (A-2) one has $\forall A \epsilon \mathbf{A}$

$$\|A\|^2 = tr(AA^T) = tr(A^2) = \sum_{i=1}^{n} \lambda_i^2(A)$$

where the $\lambda_i(A)$ are the eigenvalues of A. Since A is positive definite, these eigenvalues are all positive. Now

$$\sum_{i=1}^{n} \lambda_i(A) \le c \tag{A-7}$$

so that

$$c^2 \ge \left[\sum_{i=1}^{n} \lambda_i(A)\right]^2 > \sum_{i=1}^{n} \lambda_i^2(A) = \|A\|^2$$

or

$$\|A\| < c < \infty \tag{A-8}$$

and $\mathbf{A}$ is therefore bounded.

(2) Proof $\mathbf{A}$ is closed:

To show that $\mathbf{A}$ is closed one must show for every sequence $A_n \epsilon \mathbf{A}$ which converges to $A \epsilon \mathbf{M}$ that $A \epsilon \mathbf{A}$. Consider the inner product $\langle x, A_n y\rangle$ and look at the expression

$$\begin{aligned} \langle x, A_n y\rangle - \langle x, Ay\rangle &= \langle x, (A_n - A)y\rangle \\ &\le \|x\| \cdot \|(A_n - A)y\| \\ &\le \|A_n - A\|_\infty \|y\| \cdot \|x\| \end{aligned} \tag{A-9}$$

Eq. (A-9) implies that

$$\lim_{n\to\infty} \langle x, A_n y\rangle = \langle x, Ay\rangle \tag{A-10}$$

An identical argument shows that

$$\lim_{n\to\infty} \langle A_n x, y \rangle = \langle Ax, y \rangle \tag{A-11}$$

But $A_n = A_n^T$ so that

$$\langle A_n x, y \rangle = \langle x, A_n y \rangle ; \quad \forall n \tag{A-12}$$

Since the limit is unique (A-10), (A-11) and (A-12) together imply that

$$\langle x, Ay \rangle = \langle Ax, y \rangle \tag{A-13}$$

or

$$A = A^T \tag{A-14}$$

i.e., A is symmetric. Observe that

$$\begin{aligned} \operatorname{tr} A - \lim_{n\to\infty} \operatorname{tr} A_n &= \lim_{n\to\infty} (\operatorname{tr} A - \operatorname{tr} A_n) \\ &= \lim_{n\to\infty} \operatorname{tr} (A - A_n) \\ &= \lim_{n\to\infty} \operatorname{tr} [(A - A_n) \cdot I] \\ &= \lim_{n\to\infty} \langle (A - A_n); I \rangle \\ &\leq \lim_{n\to\infty} \|A - A_n\| \cdot \|I\| = 0 \end{aligned} \tag{A-15}$$

Thus

$$\lim_{n\to\infty} \operatorname{tr} A_n = \operatorname{tr} \lim_{n\to\infty} A_n = \operatorname{tr} A \tag{A-16}$$

Now since $A_n \epsilon \mathbf{A}$ one can write

$$\operatorname{tr} A_n \le c \tag{A-17}$$

Taking the limit of both sides of (A-17) as $n \to \infty$ and using (A-16) yields

$$\operatorname{tr} A \le c \tag{A-18}$$

Similarly

$$b \le \operatorname{tr} A \tag{A-19}$$

Finally since every term in the sequence is positive definite A must be positive definite and therefore $A \epsilon \mathbf{A}$; i.e., $\mathbf{A}$ is closed.

It should be clear that the theorem is true if $\mathbf{A}$ is the set of all real symmetric positive semi-definite matrices such that

$$0 \le b \le \operatorname{tr} A \le c < \infty; \quad \forall A \epsilon \mathbf{A}$$

The proof is identical to the above. The fact that V_Q and V_R are bounded and closed follows immediately.

BIBLIOGRAPHY

1. Kalman, R.E., "A New Approach to Linear Filtering and Prediction Theory," J. of Basic Engineering, Trans. ASME, Series D, Vol. 82, pp. 35-45, March 1960.

2. Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," J. of Basic Engineering, Trans. ASME, Series D, Vol. 83, pp. 95-108, March 1961.

3. Kumar, K. S. and Sridhar, R., "On the Identification of Linear Systems," Proceedings of 1964 Joint Automatic Control Conference, pp. 361-365.

4. Shellenbarger, J. C., "Estimation of Covariance Parameters for an Adaptive Kalman Filter," Proceedings of National Electronics Conference 1966, Vol. 22, pp. 698-702.

5. Abramson, P. D., Jr., "Simultaneous Estimation of State and Noise Statistics in Linear Dynamical Systems," (ScD. Thesis), Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, 1968.

6. Carny, T. M. and Goldwyn, R. M., "Numerical Experiments with Various Optimal Estimators, " Journal of Optimization Theory and Applications, Vol. 1, No. 2, 1967.

7. Nishimura, T., "Error Bounds on Continuous Kalman Filters and the Application to Orbit Determination Problems," IEEE Trans. on Automatic Control, Vol. AC-12, No. 3, June 1967.

8. Roher, R. A. and Sobral, M., "Sensitivity Considerations in Optimal System Design," IEEE Trans. on Automatic Control, Vol. AC-10, pp 43-48, January, 1965.

9. Dorato, P. and Kestenbaum, A., "Application of Game Theory to Sensitivity Design of Optimal Systems", IEEE Trans. on Automatic Control, Vol. AC-12, pp. 85-87, February, 1967

10. Salmon, D. M., "Minimax Controller Design," IEEE Trans on Automatic Control, Vol. AC-13, pp. 369-376, August, 1968.

11. Eggleston, H. C., Convexity, Cambridge University Press, London, 1958.

12. Karlin, S., Mathematical Methods and Theory in Games, Programming, Economics, Vol. II, Addison-Wesley, Reading, Mass., 1959.

13. Athans, M. and Schweppe, F. C., "Gradient Matrices and Matrix Calculations", MIT Lincoln Laboratory, Lexington, Mass., Tech. Note 1965-53, November, 1965.

14. Bellman, R. E., Introduction to Matrix Analysis, McGraw-Hill Book Company, Inc., New York, 1960.

15. DeRusso, P. M., Roy, R. J., Close, C. M., State Variables for Engineers, John Wiley and Sons, Inc., New York, New York, 1965.

16. Kalman, R. E., "New Methods and Results in Linear Prediction and Filtering Theory", Research Institute for Advanced Studies, Baltimore, Maryland, Tech. Report 61-1.

17. Blackburn, T. R., "Solution of the Algebraic Matrix Riccati Equation via Newton-Raphsen Iteration," Preprints of Ninth Joint Automatic Control Conference, pp. 940-945, June 1968.

18. Chen, C. F., and Shieh, L. S., "A Note on Expanding $PA+A^TP=-Q$", IEEE Trans. on Automatic Control, Vol. AC-13, No. 1, pp. 122-123, February, 1968.

19. Pinson, J. C., "Inertial Guidance for Cruise Vehicles", Chap. 4 in Guidance and Control of Aerospace Vehicles, Leondes, C. T., Ed., McGraw-Hill, New York, New York, 1963

20. Bryson, A. E., and Johansen, D. E., "Linear Filtering for Time Varying Systems Using Measurements Containing Colored Noise", IEEE Trans. on Automatic Control, Vol. AC-10, No. 1, pp. 4-10, January, 1965.

21. Athans, M. and Falb, P. L., Optimal Control, McGraw-Hill, New York, New York, 1966.

22. D'Appolito, J. A., and Hutchinson, C. E., "Low Sensitivity Filters for State Estimation in the Presence of Large Parameter Uncertainties, "IEEE Trans. on Automatic Control, Vol. AC-14, No. 3, June 1968.

23. Catlin, D. E., Lectures on Vector Algebra and Analysis for the Physical Sciences, unpublished course notes, University of Massachusetts.

24. Porter, W. A., Modern Foundations of Systems Engineering, MacMillan Co., New York, New York, 1966.